The Failure of Too Cool for School with Endogenous Effort

Andrew Clausen
clausen@econ.upenn.edu

January 15, 2007

1 Introduction

Feltovich et al. (2002) introduced the concept of countersignalling, where the brilliant signal their position by understatement. For example, top students sometimes do not work hard at achieving high grades, and rely on other signals such as recommendations. Feltovich et al argue that this strategy works in their favor in two ways. Firstly, they avoid some hard work. More importantly, they can separate themselves from mediocre students who need to get high grades by demonstrating how confident they are that the other signals will be strong enough.

However, in the environment Feltovich et al. (2002) consider, the players are exogenously endowed with their quality levels. Coate and Loury (1993) argue that signalling in contexts such as education and advertising is preceeded by an investment in quality. This paper argues that countersignalling equilibria are not robust, in the sense that countersignalling can not support incentives to invest in quality.

Most of the examples given by Feltovich et al. (2002) seem plausible because they have signals of secondary importance, in the sense that the signalling continuation games discussed are not the primary motivation for acquiring skills:

- New money flaunt their wealth with sports cars and philanthropic projects, whereas old money are above such displays. (Old money do not choose to acquire wealth – they inherit it.)
- Only minor officials need to prove their status by making petty demands. (Skill investments are not relevant.)
- Mediocre students answer as many of their teachers' questions as they can, whereas the best students feel embarrassed to prove their knowledge of trivial points. (But is impressing teachers really the primary motivation for working hard at school?)

However, in Feltovich et al. (2002)'s leading example of education signalling eluded to in the title of their paper *"Too Cool for School?"*, high grades are motivated by securing well paid employment. If education does provide useful skills, then countersignalling should be studied in a model with endogenous skill investment. Of course, if employers value pain tolerance rather than skills, an exogenous quality model is adequate.

The next section describes a signalling model with endogenous effort. Section 3 gives a benchmark example of countersignalling when effort is exogenously restricted. Section 4 then proves that there is no countersignalling equilibrium when effort is not exogenously constrained.

2 Model

This section introduces a signalling model with endogenous effort, defines the equilibrium concepts and compares the model with the original Feltovich et al. (2002) model.

Each player chooses how much effort to put into acquiring skills, and how many tests to take (1 or 2). The cost of effort differs among the players. Exerting effort has two effects: it increases the chances of acquiring skills, and increases the chances of passing the tests. Employers then offer jobs with wages based on the number of tests passed and the number of tests attempted.

- Let $C = \{c_l, c_m, c_h\} \subseteq \mathbb{R}$ be the set of possible skill investment costs. Without loss of generality, assume $c_l \leq c_m \leq c_h$.
- Let $P = \{(c, i) : c \in C, i \in [0, \lambda_c]\}$ be the set of people. λ_c represents the measure of people with skill investment cost c. With some abuse of notation, I will write $c(p) = c(c, i) = c$ is the cost of skill investment for person $p \in P$.
- Each person chooses an effort $e(p) \in [0, 1]$ and how many tests $n(p) \in \{1, 2\}$ to take.
- Let $k(e)$ be such that $k(e)c(p)$ is the cost of to a person of type c of putting in an educational effort of e. I will assume k is differentiably strictly increasing and differentiably strictly convex in e.
- Let $\phi(p) \in \mathbb{R}_+$ be the cost for person p of taking the second test. (The first test is free.) I will assume that ϕ only depends on i/λ_c and that $\phi(0) = 0$. So, there exists some function $f : \mathbb{R} \to \mathbb{R}$ such that $\phi(c, i) = f(i/\lambda_c)$.
- Each test is pass/fail. The probability of person p passing is $e(p)$.
- Person p has skills ($q(p) = 1$) with probability $e(p)$, and no skills ($q(p) = 0$) otherwise. The test(s) and the skills are IID.

  That is, skill investment increases the chance of acquiring skills, and of getting a good signal. This independence assumption means that the signal is of effort only, and does not reveal any (extra) information about whether skills were acquired.
- A person p who passed $m = m(p)$ out of $n = n(p)$ attempted tests gets a market wage of

  $$w(m, n) = \max\{\underline{\mathrm{w}}, \mathbb{P}(q(p) = 1 | m(p) = m, n(p) = n)\},$$

  where p is drawn uniformly from P. (This could be justified by a story like Fang (2001) of two firms with two technologies; one with worker productivity of $\underline{\mathrm{w}}$, and the other of 1.)
- Thus, person p's ex post utility function is

  $$u(p, e, m, n, w) = w(m, n) - k(e)c(p) - (n - 1)\phi(p),$$

  and ex ante expected utility function is

  $$U(p, e, n, w) = \mathbb{E}_m[w(m, n)] - c(p)k(e) - (n - 1)\phi(p).$$

Definition 1 (Perfect Bayes-Nash equilibrium). *A* Perfect Bayes-Nash equilibrium *is* (e^*, n^*, w^*) *such that*

- (e^*, n^*) *maximizes person* p*'s expected utility given* $(w^*, c(p), \phi(p))$.
- $w^*(m, n)$ *is obtained from Bayes' rule from* (e^*, n^*)*, as defined above.*

Definition 2 (Countersignalling equilibrium). *Consider a PBNE* (e^*, n^*, w^*). *Let* π_c^* *be the portion of people with effort cost* c *that take both tests. That is,* $\pi_c^* = \lambda(\{p \in P : c(p) = c, n^*(p) = 2\})/\lambda_c$. *A* countersignalling *equilibrium is a PBNE such that* $\pi_m^* > \pi_l^*, \pi_h^*$.

Definition 3 (Trivial equilibrium). *A PBNE* (e^*, n^*, w^*) *is trivial if* $e^*(c, \cdot) = 0$ *for some* $c \in \{c_l, c_m\}$.

Note that there is always a trivial equilibrium.

Like Feltovich et al. (2002), each player has an exogenous signal (the first test), and chooses whether to provide an additional endogenous signal (second test). Moreover, there are three types, so that the medium quality types can separate themselves from the lows by using the endogenous signal, and the high quality types can separate themselves from the mediums by pooling with the lows.

Apart from the motivating difference, that effort is chosen endogenously, there are several other modifications to the Feltovich et al. (2002) model. All signals are binary (pass/fail) and all signals (not just the exogenous signal) are stochastic from the test-taker's point of view for simplicity. Moreover, skills are binary; differences in the population arise from differences in *expected* skills. In addition, each player has an idiosyncratic cost to taking the second test that is independent of their cost of acquiring skills, like Norman (2003).

Since the endogenous signal is binary, the definition of *countersignalling equilibrium* is different. Rather than requiring all medium types to signal with higher intensity than the other types, my definition requires a larger fraction of medium types take the second test than the other types.

Finally, this model explicitly constructs the demographics of the population P. The unpublished appendix of Moro and Norman (2004) suggests this approach as a mathematical trick to resolve some paradoxes regarding the law of large numbers. Instead, I use this approach to aid intuition. I did not explicitly construct the entire measure space (I excluded the skill and test random variables), so my model is still subject to the concerns raised in Uhlig (1996). It would be conceptually trivial, but somewhat tedious, to resolve this issue by including reservation effort levels for the two tests and the skill random variables in the definition of P, as discussed in the unpublished appendix of Moro and Norman (2004).

3 Countersignalling with exogenous effort

Since the model described in the previous section differs in several respects to the original Feltovich et al. (2002) countersignalling model, this section gives an example of a countersignalling equilibrium when effort is exogenously given to the players. Of course, Feltovich et al. (2002) give a detailed discussion of the existence of countersignalling equilibria, including several formal propositions. The purpose of this section is limited to showing that the model from the previous section was not rigged to preclude countersignalling equilibria.

I exogenously set effort $e(c, n) = c$, where $c \in C = \{0.9, 0.8, 0.1\}$, the minimum wage $\underline{w} = 0.2$, and $\phi(i) = 0.05\sqrt{i}$, and finally, $\lambda_c = 1$.

I will briefly outline the algorithm I used to compute the equilibrium. The source code (written in R) is available upon request.

- The inputs are the parameters listed above, including the exogenously given effort.
- Set an initial π^0, and compute w^0 from (π^0, e) using Bayes' rule.
- Repeat the following for $t \in \{1, 2, \dots\}$, stopping when wages w^t have approximately converged (i.e. when $w^t \approx w^{t-1}$).
  - Compute π^t from (w^{t-1}, e), by finding the person for each $c \in C$ who is indifferent between taking one or two tests.

- Compute $\hat{w}^t$ from (π^t, e) using Bayes' rule.
- Compute $w^t = 0.9w^{t-1} + 0.1\hat{w}^t$. (This conservative convex combination is to avoid overshooting.)

- Output the converged (w, π).

It is straightforward to prove that the output of this algorithm is approximately an equilibrium. The intuition is as follows: the algorithm constructs the test decision π so that it is best responding to wages w, and wages are determined by the aggregate decisions π, and only stops when they are in equilibrium with each other.

The algorithm does not terminate on all inputs. Decreasing the weight of $\hat{w}^t$ – which is 0.1 in the above description – increases the set of inputs for which the algorithm converges.

Since there are multiple equilibria, the algorithm will not necessarily select a countersignalling equilibrium. Starting with a π^0 vector that has a countersignalling structure helps. Decreasing the weight of $\hat{w}^t$ makes the algorithm less likely to jump out of countersignalling territory. (The algorithm is not stochastic, but the system is somewhat chaotic.)

For the parameter values described above, $(\pi_l^*, \pi_m^*, \pi_h^*) = (0.002107122, 0.539962307, 0.222107675)$ and w^* given in the table below form a countersignalling equilibrium:

	0	1	2
1	0.2617052	0.8263219	NA
2	0.2000000	0.6688591	0.7960839

Recall that wages w^* are determined by the number of tests passed (columns) and the number of tests taken (rows).

The π^* vector in this equilibrium clearly exhibits countersignalling behaviour, as a larger measure of the medium effort group take the second test than either the high or low effort groups.

I conclude that the model from Section 2 is rich enough to support countersignalling equilibria.

4 Failure of countersignalling with endogenous effort

This section shows that there is no countersignalling equilibrium if $k(e) = e^2$. Most of the proof does not depend on this assumption, and I believe the result can be generalized substantially. I have not been able to find any countersignalling equilibria with other functional forms such as $k(e) = e^3$.

The first lemma shows that the effort-test decision can be decomposed into two decisions; once the test decision has been made, effort only depends on the cost of effort c and what test decision was made. The cost of taking the test is irrelevant to the effort decision.

Moreover, the dependence of effort on c is well-behaved, in the sense that it is continuously differentiable, decreasing, and strictly convex. Note that the domain of E is extended from C to a connected open set.

Lemma 1 (Representation of Effort)**.** *Suppose $k(e) = e^2$. If (e^*, n^*, w^*) is an equilibrium then there exists a function $E^*(c, n)$ such that $e^*(p) = E(c(p), n^*(p))$. Moreover, if $E^*(c, n) \in (0, 1)$, then $E^*(\cdot, n)$ is differentiably strictly decreasing and differentiably strictly convex in a neighbourhood of c. Finally, in a non-trivial equilibrium, $E^*(\cdot, n) > 0$.*

Proof. Recall that each person's problem is

$$\begin{aligned}\max_{e,n} U(p, e, n) &= \max_{e,n} \mathbb{E}_m[w^*(m, n)] - c(p)k(e) - (n-1)\phi(p) \\ &= \max_{e} \mathbb{E}_m[w^*(m, n^*(p))] - c(p)k(e) - (n^*(p)-1)\phi(p).\end{aligned}$$

Firstly, if $n^*(p) = 1$, then

$$\max_{e,n} U(p, e, n) = \max_{e}(1 - e)w^*(0, 1) + ew^*(1, 1) - c(p)k(e).$$

This gives the first order condition

$$w^*(1, 1) - w^*(0, 1) = c(p)k'(e).$$

The left side is the marginal benefit of increasing effort e, and the right side the marginal cost. The marginal benefit is constant, and the marginal cost is strictly increasing, since k is strictly convex. So there is at most one solution e to this FOC. If there is no *interior* solution to the optimization problem, then $e^*(p) \in \{0, 1\}$ (since there is a solution to the problem from the compactness of the choice set and continuity of the objective).

In particular, there exists a function $E^*(c, n)$ such that every p with $n^*(p) = 1$ has $e^*(p) = E(c(p, n^*(p))$. By the inverse function theorem $E^*(\cdot, n^*(p))$ is differentiable on $\{c \in \mathbb{R}_+ : E(c, 1) \in (0, 1)\}$.

Moreover, $(c, e) \mapsto w^*(1, 1) - w^*(0, 1) - ck(e)$ has the c partial derivative constant and negative, and the e partial derivative strictly decreasing and negative. So the implicit function theorem implies $E^*(\cdot, 1)$ is differentiably strictly convex and differentiably strictly decreasing.

The $n^*(p) = 2$ case is similar. The first order condition is

$$2e[w^*(2, 2) - 2w^*(1, 2) + w^*(0, 2)] + 2w^*(1, 2) - 2w^*(0, 2) = k'(e)c(p).$$

Unlike the previous case, the marginal benefit of effort is not constant, and there may be more than one intersection between marginal cost and marginal benefit. If $k(e) = e^2$, then there is a unique closed form solution,

$$\begin{aligned} 2ec &= 2e[w^*(2, 2) - 2w^*(1, 2) + w^*(0, 2)] + 2w^*(1, 2) - 2w^*(0, 2) \\ 0 &= 2e[w^*(2, 2) - 2w^*(1, 2) + w^*(0, 2) - c] + 2w^*(1, 2) - 2w^*(0, 2) \\ E^*(c, 2) &= \frac{w^*(1, 2) - w^*(0, 2)}{2w^*(1, 2) - w^*(2, 2) - w^*(0, 2) - c}. \end{aligned}$$

Finally, in a non-trivial equilibrium, the numerators of these expressions are all non-zero, so effort is non-zero. ☐

The test value $\Delta(c)$ to a type c person is the expected gain from taking the second test.

Definition 4 (Test Value)**.** *Let (e^*, n^*, w^*) be any PBNE. Let $V_n(c) = U(p, E^*(c, n), n, w^*)$ be the value of taking n tests for person $p = (c, 0)$. Let $\Delta(c) = V_2(c) - V_1(c)$ be the benefit of taking 2 tests instead of 1.*

The following lemma asserts that the person indifferent between taking one or two tests finds the cost of taking the second test equal to the expected gain.

Lemma 2 (Indifferent Test Taker)**.** $\phi(c, \pi^*_c \lambda_c) = \Delta(c)$.

Proof. Clearly, person $p = (c, \pi^*_c \lambda_c)$ is indifferent between taking one or two tests, since the cost and benefit of taking the extra test are equal. ☐

The following lemma asserts that a persons' utility varies in a well-behaved manner as the cost of acquiring skills varies.

Lemma 3 (Smoothness of Effort)**.** *If $E^*(c, n) \in (0, 1]$, then V_n is differentiable in an open neighbourhood of c with $V_n'(c) = -k(E^*(c, n))$.*

Proof. Firstly, if $E^*(c,n) \in (0,1)$, then from the representation of effort lemma, $E^*(\cdot,n)$ is differentiable at c. By the envelope theorem, interior solutions of the optimization problem have

$$V_n'(c) = U_c(c, E^*(c,n), n, w^*).$$

Substituting from the definition of U gives $V_n'(c) = -k(E^*(c,n))$.

Otherwise, if $E^*(c,n) = 1$, then $V_n'(c) = -k(1)$. Clearly, this extension of V_n' is continuous, so V is continuously differentiable. □

The next lemma shows that there is a unique cost c^* of investing in skills that maximizes the gain from taking the second test. People with this cost c^* invest the same amount in skills, regardless of how many tests they take.

Lemma 4 (Maximum Gain from Second Test). *If (e^*, n^*, w^*) is a countersignalling equilibrium, then there exists a unique local and global maximizer c^* of $\Delta(\cdot)$. Moreover, $E^*(c^*,1) = E^*(c^*,2) \in (0,1)$, with*

$$E^*(c^*,\cdot) = \frac{[w^*(1,2) - w^*(0,2)] - \frac{1}{2}[w^*(1,1) - w^*(0,1)]}{2w^*(1,2) - [w^*(2,2) + w^*(0,2)]}.$$

Proof. In a countersignalling equilibrium, $\pi_m^* > 0$, so c_m has $\Delta(c_m) > 0$. Since Δ is bounded by construction, initially constant (where the optimal effort is $e = 1$) and eventually decreases (since $\pi_m^* > \pi_h^*$), its supremum is unchanged if restricted to some sufficiently large compact set. It also continuous by construction, so by Weierstrass it must attain a maximum c^*.

From the previous lemma, $\Delta(c)$ is differentiable at c if $E^*(c,\cdot) \in (0,1]$. But a maximizing c^* must have have $E^*(c^*,\cdot) > 0$, so the maximum is a stationary point with $\Delta'(c^*) = 0$.

If $\Delta'(c^*) = -k(E^*(c^*,2)) + k(E^*(c^*,1)) = 0$, then $E^*(c^*,1) = E(c^*,2)$, as k is injective.

Let $e = E^*(c^*,1) = E^*(c^*,2)$. From the equation and the FOCs, we deduce

$$\begin{aligned}
c^*k(E^*(c^*,1)) &= c^*k(E^*(c^*,2)) \\
w^*(1,1) - w^*(0,1) &= (2e-2)w^*(0,2) + (2-4e)w^*(1,2) + 2ew^*(2,2) \\
w^*(1,1) - w^*(0,1) + 2w^*(0,2) - 2w^*(1,2) &= 2e[w^*(0,2) - 2w^*(1,2) + w^*(2,2)] \\
e &= \frac{[w^*(1,2) - w^*(0,2)] - \frac{1}{2}[w^*(1,1) - w^*(0,1)]}{2w^*(1,2) - [w^*(2,2) + w^*(0,2)]}.
\end{aligned}$$

Then, $c^* = \frac{w^*(1,1) - w^*(0,1)}{k(e)}$. Since c^* is unique, it must be the unique local and global maximum. □

Proposition 1 (Non-existence of countersignalling equilibria). *If $k(e) = e^2$, then there is no countersignalling equilibrium.*

Proof. If $k(e) = e^2$, then the $n = 2$ first order condition gives

$$\begin{aligned}
2ec &= 2e[w^*(2,2) - 2w^*(1,2) + w^*(0,2)] + 2w^*(1,2) - 2w^*(0,2) \\
0 &= 2e[w^*(2,2) - 2w^*(1,2) + w^*(0,2) - c] + 2w^*(1,2) - 2w^*(0,2) \\
E^*(c,2) &= \frac{w^*(1,2) - w^*(0,2)}{2w^*(1,2) - w^*(2,2) - w^*(0,2) - c}
\end{aligned}$$

Let c^* be the maximum of Δ, as in the previous lemma.

By the monotonicity of $E^*(\cdot, 2)$, we know $E^*(c_h, 2) < E^*(c^*, 2)$. But this is impossible:

$$\frac{w^*(1,2) - w^*(0,2)}{2w^*(1,2) - w^*(2,2) - w^*(0,2) - c_h} < \frac{[w^*(1,2) - w^*(0,2)] - \frac{1}{2}[w^*(1,1) - w^*(0,1)]}{2w^*(1,2) - [w^*(2,2) + w^*(0,2)]}$$

$$w^*(1,2) - w^*(0,2) < [w^*(1,2) - w^*(0,2)] - \frac{1}{2}[w^*(1,1) - w^*(0,1)]$$

$$0 < -\frac{1}{2}[w^*(1,1) - w^*(0,1)]$$

$$w^*(0,1) > w^*(1,1).$$

This is clearly absurd. □

5 Conclusions

In the model introduced in this paper, a countersignalling culture can not motivate people to invest in effort, and is not sustainable in equilibrium. This suggests that countersignalling can only occur either when the qualities of the sender are exogenous, or when the receiver of the (counter)signal is only of secondary importance to the sender.

A testable implication is that countersignalling only occurs either when quality is exogenously given, or when the receiver is not the primary target to be convinced. For example, brilliant students send elaborate resumes in their university applications, but do not need to announce their grades among their peers to prove themselves.

References

Coate, S. and Loury, G. (1993). Will Affirmative-Action Policies Eliminate Negative Stereotypes? *American Economic Review*, 83(5).

Feltovich, N., Harbaugh, R., and To, T. (2002). Too Cool for School? Signalling and Countersignalling. *RAND Journal of Economics*, 33(4).

Moro, A. and Norman, P. (2004). A General Equilibrium Model of Statistical Discrimination. *Journal of Economic Theory*, 114(1).

Norman, P. (2003). Statistical Discrimination and Efficiency. *Review of Economic Studies*, 70(3).

Uhlig, H. (1996). A Law of Large Numbers for Large Economies. *Economic Theory*, 8(1).