Aalto University
School of Business

SIGNALING AND COUNTERSIGNALING

Bachelor's Thesis
Aaro Häkkinen
Aalto University School of Business
Bachelor's Programme in Business
Fall 2018

Table of Contents

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of bachelor's thesis

Author Aaro Häkkinen

Title of thesis Signaling and countersignaling

Degree Bachelor of Science (Economics and Business Administration)

Degree programme Bachelor's Programme in Business

Thesis advisor(s) Mikko Mustonen

Year of approval 2018 **Number of pages** 12 **Language** English

Abstract

Signaling is relevant for markets and games with incomplete information, and it occurs when a player believes he can improve his payoff by a decision to signal. In order to study signaling games, it is meaningful to seek possible equilibria. Perfect Bayesian equilibrium is a solution concept that can be used for this. Related to strategies of players, countersignaling is a phenomenon in which a player of higher quality invests less in signals than a player of lower quality. This kind of equilibrium requires an existence of additional noisy information that can change the interpretation of signals.

Keywords signaling, countersignaling, game theory

1. Introduction

The phenomenon of signaling can be explained by comparing markets with perfect information to markets with incomplete information. A market where all parties have perfect information has no use for signaling. Both buyers and sellers have a complete understanding of all goods and services. Therefore, the perceived value of products by different parties cannot be affected by releasing information.

Signaling becomes relevant when a market, or game, has an information asymmetry. When certain information is known to only one side of the market, it can be beneficial to share this information with the other side of the market. Signaling occurs when a player believes that disclosing certain information is interpreted by other players in a way that the increase in his payoff exceeds the possible costs of signaling. In the initial papers of signaling literature, Spence (1973) presented a model to study the signaling behaviour of the markets of incomplete information.

The topic of signaling is too extensive to be completely covered in this paper. This is the case especially when it comes to different refinements of signaling equilibria. Thus, this overview attempts to summarise the framework of signaling and to link it to the concept of countersignaling. After reading this thesis, the reader should understand why signaling as well as countersignaling occur and how signaling games are studied.

In the next section, the definition of signaling is explained by using the examples of Spence's signaling model as published in 1973. The solution concepts for signaling games that were refined over the next few decades are introduced in the third section. The fourth section presents countersignaling, a phenomenon that occurs in some cases of signaling. Finally, the fifth section concludes.

2. Defining signaling

A paradigm case of a market with asymmetric information is the job market (Spence, 1973). The employer does not usually know the productive capability of a job applicant. Even after a decision to hire an individual, it may take time to see whether the employee will meet the expectations. Hence, hiring can be seen as an investment decision.

Presuming the employer is risk-neutral, wage can be assumed as an individual's marginal contribution to the firm. As mentioned in the previous paragraph, the marginal contribution of an individual cannot be predicted before hiring. Therefore, the wage is determined by how the employer perceives the value of an investment opportunity based on limited knowledge.

The employer will eventually learn the productive capability of hired individuals. This acquired knowledge is used to adjust the employer's beliefs. By combining the new knowledge and the current set of beliefs, the employer updates his beliefs related to the attributes of employees. Beliefs can be defined as a rational assessment of probabilities. They may include a probability distribution over possible outcomes as well as the types of other players. Beliefs are based on the individual's information set rather than his feelings.

Spence (1973) divides the observable attributes of a job applicant into two separate categories: indices and signals. Indices are attributes that cannot be altered by the applicant. Some of them are static, such as race and sex; while others change over time, like age. However, signals include attributes that can be affected by the applicant. Education is an example of a signal which will be used later in this paper.

While signals can be altered by the applicant, there can be costs related to making these adjustments. These costs are instinctively called signaling costs. They include, but are not limited to, money, time and effort. From the applicant's point of view, it is rational to choose signals that maximize the difference between expected income and signaling costs.

One could interpret the reasoning above as a claim that people choose to get an education only to increase their income. There certainly are more reasons to obtain a degree. Some do it to meet the expectations of their families, while others are motivated by the new experiences it offers. Implementing all motives separately into a model would be impossible – and arguably, higher income is the foremost reason to acquire an education. However, all the alternative motives can be included in the signaling costs. Although the financial costs of an education may be the same for all, differences in motivation result in varying signaling costs between individuals.

Different motives are not the only reason for the disparity of signaling costs. It has to be assumed that the productive capability of an applicant correlates positively with the applicant's ability to acquire an education. In other words, signaling costs are negatively correlated with productive capability. When this assumption holds true, it does not necessarily mean that applicants with higher productive capability should get an education. It merely means it is easier to acquire an education as an applicant with higher productive capability when compared to those with lower productive capability.

What if the critical assumption above does not hold true? First, if there were no correlation between signaling costs and productive capability of an applicant, all applicants would choose to signal in an identical manner, thus they would not be distinguishable by their signaling. Secondly, although there are cases where the level of an applicant's productive capability is lower than the productive capability of those with higher signaling costs, they are arguably in the minority. As an example, individuals with a combination of high cognitive and low noncognitive skills can find it relatively easy to obtain a degree in a theoretical subject while struggling to perform efficiently in work environments.

As evidence to support the assumption, there is a strong positive correlation between the level of income and the level of education (Tamborini, Kim, and Sakamoto, 2015). If it is assumed that acquiring an education does not significantly increase the productive capability of an individual, it demonstrates that the costs of signaling correlate negatively with productive capability. Also, it was later shown that the framework holds even if education increases one's productivity (Araujo, Gottlieb, and Moreira, 2007).

Due to the continuous nature of the job market, it can be viewed as an ongoing feedback loop. The employer offers wages based on his beliefs of an applicant's parameters – in other words, indices and signals. Applicants base their signaling decisions on offered wages in order to maximize their net return of signaling costs. After hiring applicants, the employer observes the productive capabilities of individuals with different sets of parameters and adjusts his beliefs accordingly, and the loop continues.

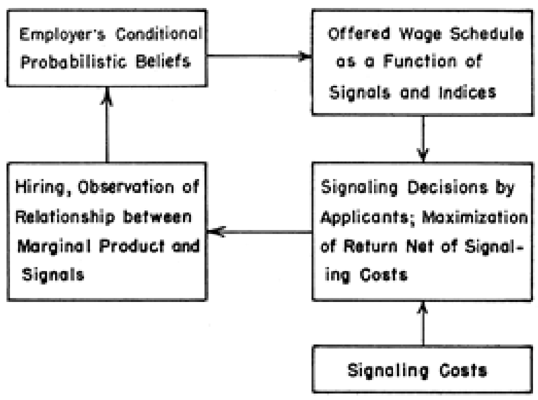

Figure I - Informational feedback loop of the job market (Spence, 1973)

In order to study this system, it is practical to seek a configuration where the loop remains unchanged. This happens when the employer's beliefs do not require adjusting between rounds because observed information matches the current set of beliefs. These beliefs that are not discarded as false, are referred to as self-confirming.

3. Signaling equilibria

According to Spence (1973), an equilibrium can be seen as "a set of components in the cycle that regenerate themselves". He approached defining properties of possible equilibria by using a simplified model that assumes two fixed wage levels, as can be seen below in Figure II.

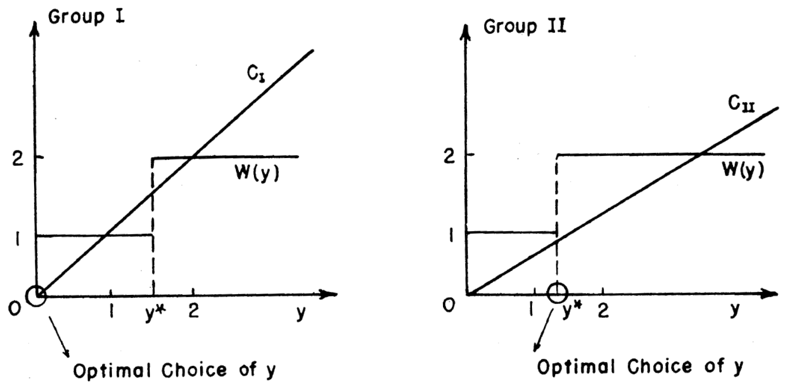

Figure II - Offered wages W and the costs of signaling C as a function of the level of education y, optimizing choice of education for two groups (Spence, 1973)

The employer defines the level of y^* which is the required level of education in order to prove higher productivity. If an individual chooses any level of education $y < y^*$, he will receive a wage of one unit. And if he chooses any level of education $y \geq y^*$, he will receive a wage of two units. Therefore, only rational levels of education are $y = 0$ and $y = y^*$.

Given the value of y^*, each individual decides between these two options based on his net return of signaling costs. In the example of Figure II, all applicants are divided into two groups based on their costs of signaling. Group I has a unit cost of y for signaling and the optimal choice for individuals in this group is $y = 0$, while Group II has a unit cost of $0.5y$ for signaling and the optimal choice for individuals in this group is $y = y^*$.

There is an infinite number of possible values for y^*, thus there is an infinite number of possible equilibria. However, the possible range of y^* that results in the same behaviour can be determined. Group I selects $y = 0$ only if $2 - y^* < 1$ while Group II selects $y = y^*$ only if $2 - 0.5y^* > 1$. Solving this simultaneous set of two equations results in $1 < y^* < 2$.

What if signaling was not possible? Each individual would be paid the average productivity of all individuals. As long as there is at least one individual who has a productivity of two, those with a productivity of one would receive a higher wage than with the possibility of signaling. Depending on the distribution of individuals with different productivities and the level of y^*, those with a productivity of two could possibly also have a higher net return. The higher the proportion of individuals with productivity of two and the closer y^* is to the upper limit of its possible range, the more they would gain if there were no possibility of signaling.

To clarify the purpose of y^*, consider that the ratio of signaling costs and the productivity difference of low and high types is altered. Neither type chooses to signal when the costs of signaling for high types exceed the difference in productivity. There is no separation between these two types. If the employer benefits from the separation, it may lower the needed level of education for y^*. As a result, obtaining an education becomes profitable for individuals of higher productivity. A similar kind of adjustment can happen both ways when the current level of y^* does not lead to separation.

In his pioneering framework, Spence (1973) left open the question of how possible equilibria should be defined. Explaining possible equilibria could be done in his simplified example. In contrast, if there is a continuum of quality levels, there is no Nash equilibrium (Riley, 1979). This overview presents perfect Bayesian equilibrium, a solution concept for dynamic games with incomplete information. There are other solution concepts that can be applied to signaling games, but they are more or less variations of perfect Bayesian equilibrium, usually with more restrictions on the beliefs. Furthermore, perfect Bayesian equilibrium has usually been preferred over other solution concepts in most applications of dynamic games with incomplete information (Fudenberg, and Tirole, 1990), but even the definition of perfect Bayesian equilibrium itself varies slightly in literature.

Perfect Bayesian equilibrium requires that strategies are sequentially rational – in other words, each player's strategy σ should maximize one's expected utility at every point of the game, given player's beliefs μ. The beliefs are updated according to Bayes' rule whenever it is applicable. Unlike many of its variations, the weakest form of perfect Bayesian equilibrium has no restrictions on the paths with zero probability also known as off-the-equilibrium paths. Revising beliefs according to Bayes' rule means that the conditional probabilities of different nodes are adjusted according to

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)},$$

where $P(A)$ and $P(B)$ are the marginal probabilities of events A and B independently, $P(A|B)$ is the conditional probability that A occurs if B has occurred and $P(B|A)$ is the conditional probability that B occurs if A has occurred. If there are paths with zero probability, Bayes' rule cannot be applied and these beliefs can be arbitrary. To be explained later in this section, perfect Bayesian equilibrium constrains the beliefs more than Bayes' rule alone.

In their works, Fudenberg and Tirole (1990, 1991) refined a formal definition of perfect Bayesian equilibrium. The following definition applies to multi-stage games with observed actions and incomplete information. Players are denoted by $i, j, \ldots, I$. Each player i has type θ_i while θ_{-i} denotes the types of his opponents, and Θ_i is a finite set of player types. The type of player is independent of others' and the probabilities of types are identical for all players. The action chosen by player i is denoted by a_i and A_i is the set of possible moves for player i. The history of moves before time period t is denoted by h^{t-1} while H^{t-1} is the set of all possible histories before time period t. The strategy σ_i of player i can be defined as a sequence of maps from

$$\Theta \times H^{t-1} \text{ to } \Delta(A_i(h^{t-1})),$$

where $\Theta \times H^{t-1}$ is the information set of player i that is used to determine his next action and $\Delta(A_i(h^{t-1}))$ is the space of probability distributions over player i's possible actions given the history of moves. Player i's payoff, or utility, u_i depends on the history including the present action h^T, his own type as well as the types of others:

$$u_i(h^T, \theta_i, \theta_{-i}).$$

The conditional probability that player i chooses action a_i depends on his own type and the history of moves:

$$\sigma_i(a_i \mid \theta_i \,, h^{t-1}).$$

Each player knows his own type, but is not given information about other players' types. As a player's actions depend on his type and those actions are observed by other players before the next time period, the history of moves alters the player's beliefs of his opponents' types. Therefore, player i's conditional beliefs of his opponents' types depend on his own type and the history of moves:

$$\mu(\theta_{-i} \mid \theta_i \,, h^{t-1}).$$

As mentioned earlier, perfect Bayesian equilibrium does not restrict off-the-equilibrium path beliefs. However, it imposes other types of restrictions on beliefs. First, all players have independent posterior beliefs, and all types of players have the same beliefs:

$$u_i(\theta_{-i} \mid \theta_i \,, h^{t-1}) \; = \; \prod_{j \neq i} u_i\big(\theta_j \;\big|\; h^{t-1}\big).$$

Secondly, the beliefs of present time period are updated to the next time period according to Bayes' rule:

$$\mu_i\big(\theta_j \;\big|\; (\; h^{t-1}, a^t\big) \; = \; \frac{\mu_i\big(\theta_j \;\big|\; h^{t-1}\big)\sigma_j\big(a_j^t \mid \theta, h^{t-1}\big)}{\sum_{\tilde{\theta}_j} \mu_i\big(\theta_j \;\big|\; h^{t-1}\big)\sigma_j\big(a_j^t \mid \tilde{\theta}_j, h^{t-1}\big)}.$$

Thirdly, if the player j chooses an action that has a conditional probability of zero, the updating of beliefs is not affected by the actions of other players:

$$\mu_i\big(\theta_j \;\big|\; h^{t-1}, a^t\big) \; = \; \mu_i\big(\theta_j \;\big|\; h^{t-1}, \hat{a}_j^t\big) \; if \; a_j^t \; = \; \hat{a}_j^t.$$

This is also known as the "no-signaling-what-you-don't-know" condition, and it is based on the earlier deduction that player $k \neq j$ cannot have any information that is not already known to player i. Lastly, most applications of perfect Bayesian equilibrium assume the beliefs are common knowledge – meaning independent players i and j should have the same beliefs of a third player's type:

$$\mu_i(\theta_k \mid h^{t-1}) \; = \; \mu_j(\theta_k \mid h^{t-1}) \; = \; \mu(\theta_k \mid h^{t-1}).$$

In order to satisfy the requirements of a perfect Bayesian equilibrium, these restrictions must hold as well as the following condition: for any time period and history of moves, the strategies of the next period are a Bayesian equilibrium of the continuation game, given the beliefs. In conclusion, perfect Bayesian equilibrium is a set of strategies and beliefs so that the beliefs are consistent with the strategies, and the strategies are optimal given the beliefs.

Possible strategies for players can be divided in three categories: pure strategies, mixed strategies and totally mixed strategies. In a pure strategy, a player's decisions are predetermined. It also means that all the other players know with certainty how this player will act. In a mixed strategy, player has a probability for each pure strategy. In this case, all the other players know the probabilities of each pure strategy but cannot tell which of these pure strategies will be chosen at random. A totally mixed strategy is a mixed strategy in which all the possible options have a positive probability.

Each equilibrium belongs to one of three categories: pooling, separating or semi-separating. In a pooling equilibrium, all types choose to signal in a similar manner, and cannot be distinguished based on their signaling. In a separating equilibrium, different types choose to signal in a distinctive manner, and therefore, can be distinguished based on their signaling. In a semi-separating equilibrium, some types choose to signal in a distinctive manner while other types choose to pool with different types. In other words, certain types can be distinguished based on their signaling while some signals can be sent by more than one type of players. Pooling and separating are both important concepts in countersignaling, which is covered in the next section.

4. Countersignaling

As described earlier in the job market signaling model, types of higher productivity may choose to invest in signaling in order to distinguish themselves from the lower types. It assumes there is no other distinctive information available to the receiver. When we add an element of noisy information, it produces new kinds of equilibria (Feltovich, Harbaugh, and To, 2002). Noisy information can be defined as information that does not accurately identify the type of a player, but either excludes or alters the probabilities of types. From the receiver's point of view, it can change the interpretation of a sender's signaling choices.

To further explain this concept, suppose there are three types of job applicants: low productivity, intermediate productivity and high productivity. In a world without noisy information, the employer would only know whether an applicant chose to obtain an education or not. In the earlier example of two types, it resulted in an equilibrium in which only high types chose to signal in order to separate themselves from low types. In this example of three types, it could result in an equilibrium in which both intermediate and high types would choose to signal while low types would not.

When noisy information is added to the equation, high types could choose not to signal which would put them in the same pool with low types. It would separate them from intermediate types that are afraid to be mistaken as a low type, and therefore, avoid pooling with low types. In contrast, high types are confident they cannot be mistaken as a low type due to differences in noisy information. Not only do they separate from the intermediate types, they also save the costs of signaling.

A perfect Bayesian equilibrium can be defined as signaling equilibrium if signaling strictly increases with a sender's type, and as countersignaling equilibrium if signaling is strictly non-monotonic (Feltovich, Harbaugh, and To, 2002). Based on these definitions, any equilibrium can always be classified as either signaling equilibrium or countersignaling equilibrium, but never as both.

There are a few requirements for countersignaling equilibria to exist. First, there has to be at least three types of quality. Consider a game with two types of quality. Noisy information would either separate both types completely or not at all. When completely separated, neither of the types would benefit from signaling. If there was no separation, the possible equilibria would be similar to the original job market example. Hence, there is no pure strategy countersignaling equilibrium with only two types of quality.

Secondly, low and intermediate types have to be insufficiently separated while low and high types have to be sufficiently separated by the additional noisy information. Only then intermediate types will seek a signal that low types do not want to imitate but does not cost more than the additional gain. And if high types were not already separated from low types, they would not be able to benefit by pooling with low types. In the real world, a casual mid-tier restaurant can be mistaken as a fast food restaurant. As a result, casual restaurants have a motive to separate themselves from the fast food chains. Fine dining restaurants do not face this same issue, so their signaling is very different and often minimal when compared to mid-tier restaurants.

When there is a range of possible signals, high types cannot simply choose a less expensive signal than intermediate types. As noted by Feltovich, Harbaugh, and To (2002), only by choosing the same signal as low types, high types can successfully discourage the intermediate types from imitating their behaviour. They also point out that an increase in the costs of signaling can encourage high types to signal. As the increase in costs affects all types, it can become too difficult for intermediate types to mimic the signaling of high types. In contrast, a decrease in the costs of signaling can help intermediate types to mimic the signaling of high types, and therefore, it motivates high types to countersignal.

Lastly, the sender cannot know the realised value of noisy information. If both sender and receiver know the full content of noisy information prior to signaling decision, then the model does not differ from the standard signaling model and countersignaling equilibria cannot exist. This assumption does not rule out the possibility of splitting up the noisy information, and thus a sender can know some parts of the noisy information as long as it remains partly unknown to the sender.

Countersignaling does not limit to a maximum of three quality types. An addition of a fourth type can result in a so-called counter-countersignaling equilibrium. Assume that these type of job applicants have the highest productivity, and for them, acquiring an education is significantly easier when compared any other type. If the low and high types do not obtain an education, the highest type may choose to pool with the intermediate types even if there is an option to send a more expensive signal. In a similar fashion as explained earlier, pooling with intermediate types helps to discourage the high types from mimicking the highest types.

Possible countersignaling equilibria does not limit to pure strategies either. In case of a mixed strategy, it is natural to assume that each player of a type with a mixed strategy does an individual decision based on the probability distribution of pure strategies. This means there are possible equilibria in which some individuals choose to signal while others of the same type choose to countersignal.

Araujo, Gottlieb, and Moreira (2007) proved that job market countersignaling occurs only if there is a difference between school and workplace technologies. As the distance between these two technologies grows, so does the likelihood of countersignaling. A perfect similarity is unlikely in the real world, but this deduction can be applied to various cases in order to understand them better.

So far it has been assumed that signaling, like acquiring an education, always has its costs. If the signaling choice was not binary, sending a higher signal would cost more than sending a lower signal. In some cases, the costs of signaling are more complicated. Cho, and Kreps (1987) created a two-player example known as Beer-Quiche game in which the costs of signaling vary between two signaling options and two types of players. In contrast to the signaling choice of education, there is no choice that is costless to all types. If a player's type is surly, he loses one unit of payoff only if he chooses to have quiche for breakfast. In an opposite manner, if a player's type is wimp, he loses one unit only if he chooses to have beer. In other words, different signals are preferred by different types which reflects on the costs.

Some signaling games do not have signaling costs at all. Bederson et al. (2018) studied costless signaling of restaurants. In Maricopa County, Arizona, every restaurant is subject to unscheduled food safety inspections twice a year. In 2011, a new voluntary letter grading was adopted. Restaurants were asked whether they would like to receive a grade ranging from A to D which would be published online. However, regardless of the restaurant's decision, the detailed inspection results as well as detailed metric for calculating letter grades were available online to the public.

The study showed that only 58 percent of restaurants chose to disclose their letter grades. The rate of disclosure was positively correlated with the grades. Restaurants with the highest grade of A were most likely to disclose their grade, while those with the lowest grade of D were least likely to opt for the letter grade. However, the upper range of grade A was less likely to disclose than the lower range of grade A. The same phenomenon could be seen within other grades as well. 49 percent of non-disclosing restaurants would have been graded as A.

Bederson et al. (2018) interpreted this as a mixture of signaling and countersignaling. Countersignaling did not occur consistently between different grades, but within the same grade restaurants. As an explanation as to why such a significant proportion of restaurants chose to hide their results, they suggested there may be a fear that today's disclosure implies a liability to continue disclosing in the future (Grubb, 2011).

5. Conclusions

The initial concept of signaling in the job market (Spence, 1973) is still valid at present, but the signaling theory has kept refining since. The development has led to new solution concepts for signaling games and ideas for equilibrium strategies. Most recent studies have shown the existence of these theories in the real life – it has proven that the signaling model can be applied to a variety of practical cases.

For a more in-depth overview, this subject could have been extended to various types of equilibrium refinements. The solution concepts of signaling games is a well-studied field that contains plenty of different perspectives. Also, a solution concept could have been applied to illustrate the occurrence of countersignaling. Rather than having difficulties with finding enough content, it was challenging to narrow the content down to the most essential topics.

To conclude this paper, consider a following example. A number of students have finished their bachelor's theses. After their extensive and less extensive research, they can be divided in three categories based on their level of knowledge: low, intermediate and high. Based on the quality of their bachelor's theses, an evaluator will estimate the type of each student and the given grades will be based on these estimations. However, there is a slight chance the quality of work does not accurately reflect the level of knowledge. Both low and high types can be mistaken as intermediate, and intermediate types can be mistaken as both low and high.

In general, students think that the length of thesis is an indicator of knowledge. Before returning their theses, students have an option to add extra length to their texts by increasing the line spacing. Low types choose to increase the line spacing, because they believe it increases their odds of being mistaken as an intermediate type. Intermediate types are anxious about being mistaken as a low type. If they do not add the extra length, their number of pages is similar to those of low types. Intermediate types also choose to increase the line spacing. However, high types are confident their knowledge cannot be mistaken as low, but they believe they can still be mistaken as an intermediate type. Do high types increase their line spacing?

References

Araujo, Aloisio, Daniel Gottlieb, and Humberto Moreira. 2007. "A Model of Mixed Signals with Applications to Countersignalling." *RAND Journal of Economics 38:* 1020-1043.

Bederson, Benjamin B., Ginger Zhe Jin, Phillip Leslie, Alexander J. Quinn, and Ben Zou. 2018. "Incomplete Disclosure: Evidence of Signaling and Countersignaling." *American Economic Journal: Microeconomics 10:* 41-66.

Cho, In-Koo, and David M. Kreps. 1987. "Signaling Games and Stable Equilibria." *Quarterly Journal of Economics 102:* 179-221.

Feltovich, Nick, Richmond Harbaugh, and Ted To. 2002. "Too Cool for School? Signalling and Countersignalling." *RAND Journal of Economics 33:* 630-649.

Fudenberg, Drew and Jean Tirole. 1990. "Perfect Bayesian Equilibrium and Sequential Equilibrium." *Journal of Economic Theory 53:* 236-260.

Fudenberg, Drew and Jean Tirole. 1991. *Game Theory Cambridge, Mass. and London: MIT Press:* 319-331.

Grubb, Michael D. 2011. "Developing a Reputation for Reticence." *Journal of Economics and Management Strategy 20:* 225-268.

Riley, John G. 1979. "Informational Equilibrium." *Econometrica 47:* 331-359.

Spence, A. M. 1973. "Job Market Signaling." *Quarterly Journal of Economics 87:* 355-374.

Tamborini, Christopher R., Changhwan Kim, and Arthur Sakamoto. 2015. "Education and Lifetime Earnings in the United States." *Demography 52:* 1383-1407.