Crime and punishment in classroom: a game-theoretic approach for student cheating

Marcelo de C. Griebeler*

Federal University of Rio Grande do Sul

April 11, 2016

Abstract

We provide the microeconomic foundations of cheating in classroom through a static game with complete information. Our setting is composed by two students, who must choose whether or not to cheat, and a professor, who must choose how much effort to exert in trying to catch dishonest students. Our findings support the determinants of cheating found by the empirical literature, mainly those related to the penalty’s level. It is also emphasized the importance of professors being well-motivated (with low disutility of effort) and worried about fairness in classroom. The two extensions of the baseline model reinforce the importance of the cost-benefit analysis to understand dishonest behavior in classroom.

Keywords: student cheating; game theory; academic dishonesty.

JEL classification: C70; D01; I21.

1 Introduction

Academic dishonesty is a serious and widespread problem in the world. Although this practice may be found in institutions of all levels of education, it is better documented in colleges and universities. A recent survey conducted in the UK found that nearly 50,000 university students have been caught cheating from 2012 to 2015. The same data show non-EU scholars are the most likely to commit the offense, which suggests the student cheating is not restricted to a specific country (The Guardian, 2016). In fact, despite

*Corresponding author. Address: Universidade Federal do Rio Grande do Sul - UFRGS, Faculdade de Ciências Econômicas, Departamento Economia e Relações Internacionais, Avenida João Pessoa 52, Centro Histórico, Porto Alegre-RS, Brazil. ZIP code: 90040-000 - Phone: 55 (51) 3308-3324. E-mail: marcelo.griebeler@ufrgs.br.
the absence of reliable data for regions such as Latin America, some studies have used alternative measures to estimate that violations of academic integrity have indeed risen in Latin American countries over the past two decades (García-Villegasa et al., 2015). A substantial rise in student cheating practices has been found in colleges and universities in the United States as well. McCabe et al. (2001), for instance, reports that the number of students that admit to engage in serious test cheating (e.g. copying from another student on the exam) increased from 39% in 1963 to 64% in 1993.

In this paper we propose a three-players static game with complete information in order to model the strategic relationship underlying the student’s individual decision of cheating in classroom. Our setting is composed by two students and a professor, who must choose how much effort to exert in trying to catch dishonest students. In the baseline model, with two identical students, our findings highlight the role of the probability of the professor detects dishonesty in driving the student’s decision. For instance, an equilibrium in which both students choose not to cheat requires that the probability of being caught committing the offense be large enough. This in turn requires a large level of professor’s effort, which is mainly determined by his disutility of effort and the relative weight given to a fair classroom - without cheating - in his utility.

We also provide two extensions of the baseline model. First, we consider the case when there exists a further punishment for the dishonest student, that is, if the student is caught cheating, he is punished by losing a constant level of utility - due to failing grade in the course, suspension or expulsion, for example - in addition to zero grade in the exam. We find that a harder punishment decreases the minimum probability required to students choose not to cheat. In fact, given the probability of being caught, there exists a level of punishment that makes not to cheat be a dominant strategy for both students. This results resembles those of the classical analysis of Economics of Crime (Becker, 1968), in which the probability of being caught and the magnitude of the punishment drive the incentives of potential offenders.

The second extension allows students to be heterogeneous. They can be heterogeneous in terms of disutility of effort, for instance. This implies they choose different levels of effort and thus receive different grades. Different grades in turn implies different potential benefits for the cheater: the student with the lower grade has more to gain by cheating than the one with higher grade. This can be seen through the fact that the minimum probability that induces the student to play fair is decreasing in his own grade and increasing in the grade of the another student. Once again our model shows a feature of Economics of Crime, namely the higher the potential benefit of the offense, the more prone to commit it the individual is. In fact, the difference in terms of grades may be so large that playing not to cheat may be a dominant strategy for the student with higher grade regardless the probability of being caught cheating.

Although our framework presents similarities with the seminal model proposed by
Gary Becker, there is an important distinction between them. In our cheating game each student is at the same time a potential offender and a potential victim. For instance, the set of potential outcomes includes one in which both students choose to cheat, case in which they both copy the exam of the another student and have his own copied by the another. On the contrary, in the classical version of Becker’s model (Becker, 1968) there is no active role for victims. Given that his framework is not a game, the only agent to play is the potential criminal. Even police or law enforcer has no role in his model. In our game, however, their role is played by the professor, which makes the probability of the cheater being caught be endogenous.

Our contribution to the literature is to provide a theoretical framework that is able to capture all the strategic features of students’ choice of cheating, and professor’s choice of how much effort to exert in order to catch cheaters. To the best of our knowledge, the vast majority of literature on academic dishonesty adopt a psychological approach to investigate the determinants of student cheating (Macfarlane et al., 2014; McCabe et al., 2001). On the one hand, individual factors such as gender, grade point average (GPA), work ethic, competitive achievement striving and self-esteem have been found having significantly influence the prevalence of cheating (Baird Jr, 1980; Ward and Beck, 1990). On the other hand, contextual factors such as faculty response to cheating, sanction threats, social learning, and honor codes have also been shown to influence dishonest behavior (Michaels and Miethe, 1989). In fact, even those studies which perform economic cost benefit analysis often do it empirically, without a microeconomic model to support their results (Bisping et al., 2008; Bunn et al., 1992).

An important exception in the above trend is the study of Briggs et al. (2013), which uses game theory to analyze collaboration in academic cheating. The authors provide a relevant discussion about the use of mathematical utility modeling - and thus game theory - with respect to ethics. Based on the reasoning developed in works such as Gibson (2003), they argued that incorporation of games such as the Prisoner’s Dilemma into the ethics issues - and thus in cheating as well - may be useful to better understand costs and benefits involved. However, their model focuses in collaboration in take-home tasks rather than in-class activities, and thus is substantially different from the framework we develop in this paper.

The game-theoretic approach we employ allows us to provide both positive and normative conclusions. First, our model fits several stylized facts found by empirical studies, such as the effect of higher penalties in decreasing prevalence of cheating, the inverse

---

1Bunn et al. (1992) is the first study to present similarities and differences between cheating and the crime of theft following Becker’s approach. The two main differences found by its authors are: a professor has a greater variety of devices to affect the costs of cheating than a police officer (e.g. disperse the class during examinations, reducing density and increasing the costs of cheating) and; exam answers have a public good dimension, such that unlike “the watch stolen from its owner, answers on exams are not taken from the owner, but only copied” (Bunn et al., 1992, p. 199). This second characteristic creates the free-riding problem in classroom.
relationship between GPA and cheating behavior, and the importance of peer cheating behavior in explaining it (McCabe et al., 2001). Second, it also provides insights that can help reduce cheating on campuses. Some of them had already been found effective by literature, such as harsh penalties imposed by both the institution and the professor. Others, however, have not received much attention, including hiring high effort professors, who value fairness in classroom. As we show below, a necessary condition for the existence of a virtuous equilibrium (without cheating) is that the professor do not be lenient.

The rest of the paper is organized as follows. In the following section we present our baseline model, composed by two identical students and a professor. We discuss the incentives each one faces and describe their processes of choice. This section also establishes necessary and sufficient conditions for the existence of a Nash equilibrium without cheating. Section 3 extends the model in two directions. First, we discuss the effects of further punishments on cheaters students, and then we allow students to be heterogeneous. Section 4 concludes and suggests some extensions. The proofs of propositions omitted in the text are shown in the appendix A.

2 The student cheating game

2.1 The baseline model

Our baseline model is composed by two identical students, A and B, and one professor (or teacher). There will be an exam in the course that the professor is in charge. Each student must choose his level of effort in studying and the professor must choose his effort to detect and punish in-class cheating. Whenever a student chooses to cheat, he does not study, such that his level of effort is equal to zero. All these actions are chosen before the exam happens, there is no communication among the players and the information is complete, such that we can model this strategic situation as a three-players static game.

We consider only one type of academic dishonesty, namely the action of one student of copying from the another student’s exam without his consent or knowledge. Therefore, we rule out common cheating practices such as helping someone on the exam and using a crib note. We also do not consider academic cheating in take-home tasks, such as representing someone else’s work as your own (e.g. sharing another’s work, purchasing a term paper or test questions in advance, paying another to do the work for you). Given this assumption, whenever the professor detects a dishonesty action, he can punish only the student that copied the exam. We assume the punishment sets the dishonest student’s grade equal to zero. If the student succeeds in cheating, his grade is equal to that of the another student.

The student’s utility is a $C^2$ function and is given by $U_i(N_i,e_i)$, where $N_i$ is his grade
on the exam and \( e_i \in [0, +\infty) \) is his level of effort in studying, with \( i = A, B \). We assume the marginal utility of the grade is positive, \( \partial U_i / \partial N_i > 0 \), and the marginal utility of the effort is negative, \( \partial U_i / \partial e_i < 0 \). Given the possibility of cheating, the grade of the student \( i \) depends on his effort, the another student’s grade, the probability of being caught cheating \( p \in [0, 1] \), and mainly on his chosen strategy, whether cheating (C) or “playing fair” (PF). Moreover, given that players are identical, they have the same utility as well as the same grade function. There are four possible cases to consider:

(i) Both students \( A \) and \( B \) choose to cheat: as in this case none of them exerts any effort in studying, both their grades are equal to zero, \( N_A = N_B = 0 \).

(ii) Both students \( A \) and \( B \) choose to “play fair”: in this case each player exerts his optimal level of effort \( e_i^* > 0 \), such that the grades are \( N_A(e_A^*) \) and \( N_B(e_B^*) \). Given the assumption of identical players, \( N_A(e_A^*) = N_B(e_B^*) \).

(iii) Student \( A \) plays fair while student \( B \) cheats: the grade of the student \( A \) is \( N_A(e_A^*) \) while the expected grade of the student \( B \) is \( N_B = N_A(e_A^*) (1 - p) \).

(iv) Student \( B \) plays fair while student \( A \) cheats: here we have the opposite of the case (iii), such that \( N_A = N_B(e_B^*)(1 - p) \) and \( N_B(e_B^*) \).

Student’s grade is increasing in his effort in studying. However, the return of the effort is decreasing. We also assume some other conditions on the behavior of this function, which may be seen equal to the Inada conditions. The assumption below summarizes and formalizes the features of the grade function.

**Assumption 2.1** The student’s grade is a \( C^2 \) function of his own effort \( e_i \), given by \( N_i : [0, \infty) \to [0, 10] \), and satisfies the following properties: \( N'_i(e_i) > 0 \), \( N''_i(e_i) < 0 \), \( N_i(0) = 0 \), \( \lim_{e_i \to -\infty} N'_i(e_i) = 0 \) and \( \lim_{e_i \to 0} N'(e_i) = +\infty \).

Whenever student \( i \) chooses to play fair, he must maximize his utility by choosing the optimal level of effort \( e_i^* \). The first order condition of his problem is then given by

\[
\frac{dU_i}{de_i} = \frac{\partial U_i}{\partial N_i} N'_i + \frac{\partial U_i}{\partial e_i} = 0, \quad (2.1)
\]

which can be understood as the equality of the marginal benefit of the effort, through the increase in the student’s grade, and its marginal disutility. We discuss the existence of such an optimal choice below.

**Proposition 2.2** Suppose that the student’s utility function has the following further characteristics: (i) \( dU_i(0, 0)/de_i > 0 \); (ii) \( \partial^2 U_i / \partial e_i^2 < 0 \), \( \partial^2 U_i / (\partial N_i \partial e_i) \leq 0 \) and \( \partial^2 U_i / \partial N_i^2 < 0 \) and; (iii) \( \lim_{e_i \to -\infty} U_i / \partial N_i = +\infty \) and \( \lim_{e_i \to +\infty} \partial U_i / \partial e_i = -\infty \). Then the first order condition of the student’s problem (2.1) has an unique global maximizer at some interior point \( e_i^* \).
The first assumption of the above result means that the total marginal effect of the effort is positive when \( e_i = 0 \). This is equivalent to make the assumption that 
\[-\partial U_i(0,0)/\partial e_i < \partial U_i(0,0)/\partial N_i \cdot N_i'(0),\]
that is, the marginal gain of utility due to the increase in the grade is higher than the disutility of effort when the level of effort is zero. Thus, our model rules out “very lazy” students. The proposition also assumes that both the disutility of effort and the marginal utility of the student’s grade increase at increasing rates for all levels of effort, 
\[\partial^2 U_i/\partial e_i^2 < 0\] and \[\partial^2 U_i/\partial N_i^2 < 0,\]
respectively. The remainder assumptions are standard and have technical roles.

Some comparative statics results may help us to understand the student’s behavior.

**Proposition 2.3** The student’s optimal level of effort is a function that: (i) is decreasing in the marginal disutility of effort; (ii) is increasing in the marginal utility of his own grade and; (iii) is increasing in the return of the effort on higher grades.

The professor’s utility depends on the grades of each student, the probability of catching students cheating in class, and his effort to catch in-class cheating \( \theta \in [0, +\infty) \). We model it as a \( C^2 \) function given by \( W(N_A, N_B, p, \theta) \). We assume the professor gets more satisfaction as students’ grades increase, that is, \( \partial W/\partial N_A = \partial W/\partial N_B > 0 \). There is also an disutility of effort, such that \( \partial W/\partial \theta < 0 \). Finally, the professor wishes the fairest possible class, which means the marginal utility of the probability of catching any student cheating is positive, \( \partial W/\partial p > 0 \). We must impose some further regularities in the professor behavior.

**Assumption 2.4** The professor’s utility function has the following further characteristics: (i) it is strictly concave for all levels of effort, that is, \( d^2 W/d\theta^2 < 0 \) for \( \theta \in [0, \infty) \) and; (ii) \( \lim_{\theta \to +\infty} dW/d\theta = -\infty \).

The characteristics of the probability function are quite standard and satisfy the Inada conditions, as we highlight below.

**Assumption 2.5** The probability of catching any student cheating is a \( C^2 \) function of the professor’s effort \( \theta \), given by \( p: [0, \infty) \to [0,1] \), and satisfies the following properties: \( p'(\theta) > 0, p''(\theta) < 0, p(0) = 0, \lim_{\theta \to +\infty} p'(\theta) = 0 \) and \( \lim_{\theta \to 0} p'(\theta) = +\infty \).

We consider two types of professor, depending on the relative magnitude of his disutility of effort. The lenient professor is characterized by a very high disutility - or a very low marginal utility from the increase in the fairness of the class - when his level of effort is zero, such that \( dW^L(a,b,0,0)/d\theta \leq 0 \) for all constant \( a, b \in [0,10] \). This means that any effort to catch dishonest behaviors does not leave the lenient professor better off, regardless students’ grades. In other words, for this type of professor the benefit from the
increase in the probability of catching is not higher than the drawback from the effort.\footnote{Literature has found evidence that the prevalence of this type of professor is not negligible. McCabe et al. (2001), for example, reports that transgressions in classroom are often overlooked or treated lightly by professors who do not want to become involved in bureaucratic procedures designed to adjudicate allegation of academic dishonesty. On the student-professor relationship and its effects on cheating in classroom see also Stearns (2001).} Observe that increases in professor’s effort - and thus in probability of catching - do not positively impact the grades for any students’ choices, so we can disregard such an effect in this case.

The another type is the \textit{severe} professor, who is characterized by $dW^S(a, b, 0, 0)/d\theta > 0$ for all constant $a, b \in [0, 10]$. Now an initial effort is worth, because the marginal benefit of increasing the probability $p$ is higher than the disutility caused by such an effort. However, this case presents a further complexity, such that we may have to consider the effect that the higher probability has on the students’ grades. For example, if only one of the students cheats, say student $A$, we have $a = N_B(e_B^*) (1 - p)$ and $b = N_B(e_B^*)$, and increases in $p$ make $a$ decrease. The impact on the severe professor’s utility is then $-\partial W^S/\partial N_A \cdot N_B p < 0$. Later we must analyze whether such an effect is large enough to overcome the other two and thus makes the severe professor mimic the lenient’s choice.

As usual, the severe professor chooses his optimal level of effort $\theta^*$ by maximizing his utility. The first order condition of his problem when student $A$ cheats and student $B$ plays fair is given by

$$\frac{dW^S}{d\theta} = p' \left( \frac{\partial W^S}{\partial p} - \frac{\partial W^S}{\partial N_A} N_B \right) + \frac{\partial W^S}{\partial \theta} = 0. \quad (2.2)$$

While marginal benefit of the professor’s effort is only $\partial W^S/\partial p \cdot p' > 0$, the marginal cost is composed by the direct disutility of effort, $\partial W^S/\partial \theta < 0$, and the potential impact on the student $A$’s grade, $-\partial W^S/\partial N_A p' N_B < 0$. When student $A$ plays fair and student $B$ cheats, the professor’s FOC is similar to (2.2), except by the exchange of subscripts. For the other two cases (both students cheat and both students play fair), the FOC is also similar to (2.2), but now without the effect on the grades.

**Proposition 2.6** Suppose that the severe professor’s utility function satisfies assumption 2.4. Then the first order condition of his problem (2.2) has an unique global maximizer at some interior point $\theta^* > 0$.

The assumptions assumed in order to assure the existence of the global maximizer above are quite standard, as we have already discussed. We can now establish comparative statics results for the professor’s optimal choice.

**Proposition 2.7** The severe professor’s optimal level of effort is a function that: (i) is decreasing in the marginal disutility of effort; (ii) is decreasing in the marginal utility of
students’ grade; (iii) is increasing in marginal utility of the probability of catching students cheating and; (iv) in increasing in the marginal return of effort on the probability.

The above result states that when marginal benefits or marginal costs change, the optimal choice changes as well. While items (i) and (ii) are related to marginal costs, items (iii) and (iv) are associated to marginal benefits. This explains why in the former the relationship is inverse and in the latter it is direct. Item (ii) deserves some attention. Observe that the students’ grades affect the professor’s level of effort only if one of them chooses to cheat. In this case, there is a negative effect: higher effort implies higher probability, which in turn decreases the cheater’s grade. Therefore, the optimal level of effort decreases when there is increases in $\frac{\partial W}{\partial N_i}$, with $i = 1, 2$, because now the negative impact of increases in the probability of catching on the grades is higher.

We can sum up the game in the two payoff matrices below. For the sake of simplicity we henceforth set $U(0,0) = 0$ and $W(0,0,0,0) = 0$.

**Professor chooses $\theta^* = 0$**

<table>
<thead>
<tr>
<th></th>
<th>Student B cheat</th>
<th>Student B play fair</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student A cheat</td>
<td>$U_A(0,0)$, $U_B(0,0)$, $W(0,0,0,0)$</td>
<td>$U_A(N_B(e^<em>_B),0)$, $U_B(N_B(e^</em>_B),e^<em>_B)$, $W(N_B(e^</em>_B),N_B(e^*_B),0,0)$</td>
</tr>
<tr>
<td>play fair</td>
<td>$U_A(N_A(e^<em>_A),e^</em>_A)$, $U_B(N_A(e^<em>_A),0)$, $W(N_A(e^</em>_A),N_A(e^*_A),0,0)$</td>
<td>$U_A(N_A(e^<em>_A),e^</em>_A)$, $U_B(N_B(e^<em>_B),e^</em>_B)$, $W(N_A(e^<em>_A),N_B(e^</em>_B),0,0)$</td>
</tr>
</tbody>
</table>

**Professor chooses $\theta^* > 0$**

<table>
<thead>
<tr>
<th></th>
<th>Student B cheat</th>
<th>Student B play fair</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student A cheat</td>
<td>$U_A(0,0)$, $U_B(0,0)$, $W(0,0,p^<em>,\theta^</em>)$</td>
<td>$U_A(N_B(e^<em>_B)(1-p^</em>),0)$, $U_B(N_B(e^<em>_B),e^</em>_B)$, $W(N_B(e^<em>_B)(1-p^</em>),N_B(e^<em>_B),p^</em>,\theta^*)$</td>
</tr>
<tr>
<td>play fair</td>
<td>$U_A(N_A(e^<em>_A),e^</em>_A)$, $U_B(N_A(e^<em>_A)(1-p^</em>),0)$, $W(N_A(e^<em>_A),N_A(e^</em>_A)(1-p^<em>),p^</em>,\theta^*)$</td>
<td>$U_A(N_A(e^<em>_A),e^</em>_A)$, $U_B(N_B(e^<em>_B),e^</em>_B)$, $W(N_A(e^<em>_A),N_B(e^</em>_B),p^<em>,\theta^</em>)$</td>
</tr>
</tbody>
</table>
2.1.1 The student’s decision

Let us consider the decision of student $A$. As we are assuming symmetry, the same results are valid for student $B$. First, suppose that the professor chooses $\theta^* = 0$ and student $B$ chooses to cheat. In this case, student $A$ chooses to play fair, because

$$U_A(N_A(e_A^*), e_A^*) > 0 = U_A(0,0).$$  \hspace{2cm} (2.3)

If the professor chooses $\theta^* = 0$ and student $B$ chooses to play fair, student $A$ now is better off by choosing to cheat, because

$$U_A(N_A(e_A^*), e_A^*) < U_A(N_B(e_B^*), 0) = U_A(N_A(e_A^*), 0),$$  \hspace{2cm} (2.4)

where we once again use the symmetry assumption.

Suppose now that the professor chooses $\theta^* > 0$. If student $B$ chooses to cheat, student $A$ chooses to play fair, because his payoffs are the same as those given by (2.3). However, if student $B$ chooses to play fair, the choice of student $A$ depend on the probability of being caught cheating. In fact, student $A$ plays fair if and only if

$$U_A(N_A(e_A^*), e_A^*) \geq U_A(N_B(e_B^*)(1-p^*), 0).$$  \hspace{2cm} (2.5)

The next proposition helps us to understand the student’s behavior.

**Proposition 2.8** There exists a probability of being caught cheating $p_{\text{min}} \in (0,1)$ such that

$$U_i(N_i(e_i^*), e_i^*) = U_i(N_j(e_j^*)(1-p_{\text{min}}), 0),$$  \hspace{2cm} (2.6)

for $i,j = A,B$ and $i \neq j$.

As the proof of the proposition - see appendix A.1 - shows, function $U_A(N_A(e_A^*), e_A^*) - U_A(N_B(e_B^*)(1-p^*), 0)$ is increasing in $p$: the higher the probability of being caught, the stronger the incentive to play fair. Thus, the above result implies playing fair is a strictly dominant strategy for student $A$ if and only if $p^* > p_{\text{min}}$. Recall that we are assuming identical students, such that the same reasoning can be used to show that student $B$ chooses to play fair regardless the another student’s choice if and only if $p^* > p_{\text{min}}$. Yet, when $p^* = p_{\text{min}}$ both students are indifferent between playing fair and cheating.

We can sum up the students’ best choices with the support of the above two payoff matrices. For, let us disregard the professor’s decision for a while. In the first matrix, given $\theta = 0$, there are two equilibria: student $A$ plays fair and student $B$ cheats and; student $A$ cheats and student $B$ plays fair. In the second one, given $\theta^* > 0$, there are three cases to consider. First, if $p^* < p_{\text{min}}$, then we have the same two equilibria found in matrix $\theta = 0$. Second, if $p^* > p_{\text{min}}$, both students choose to play fair. Third, if $p^* = p_{\text{min}}$, ...
then there are three equilibria: both choose to play fair; student A plays fair and student B cheats and; student A cheats and student B plays fair. We must now study the professor’s best choice.

### 2.1.2 The professor’s decision

Let us start with the lenient one. One can readily see that this type has a dominant strategy, namely $\theta^* = 0$. The assumption that his disutility of effort is relatively very high implies $W^L(a,b,0,0) > W^L(a,b,p^*,\theta^*)$ for any $a, b \in [0,1]$. As $\partial W/\partial N_i > 0$, it also implies $W^L(a,a,0,0) > W^L(a,a(1-p^*),p^*,\theta^*)$. This covers all possibilities.

Despite the larger complexity of the behavior of the severe professor, there are two straightforward cases. First, suppose that both students choose to cheat. In this case we have $W^S(0,0,0,0) = 0 < W^S(0,0,p^*,\theta^*)$, such that his best choice is $\theta^* > 0$. When both students choose to play fair, his best choice is once more $\theta^* > 0$, because $W^S(N_A(e_A^*),N_B(e_B^*),0,0) < W^S(N_A(e_A^*),N_B(e_B^*),p^*,\theta^*)$, which is implied by the assumption that defines the severe type.

Suppose now that student A cheats and student B plays fair. The severe professor’s best choice is $\theta^* > 0$ if and only if

$$W^S(N_B(e_B^})(1-p^*),N_B(e_B^*),p^*,\theta^*) \geq W^S(N_B(e_B^*),N_B(e_B^*),0,0).$$

(2.7)

Recall that the definition of severe professor states that $p' \cdot \partial W^S/\partial p + \partial W^S/\partial \theta > 0$ when $\theta = 0$, which means that he makes a positive effort whenever it has no effect on the students’ grades. Thus, we must check whether the marginal effect of the effort on the grades, namely their decrease as a result of the increase in the probability of catching dishonest students, is enough to overcome the benefit measured by the above derivative. Formally, the severe professor’s best choice is $\theta^* > 0$ if and only if

$$\left(p'\cdot \partial W^S/\partial p + \partial W^S/\partial \theta \right)_{\theta=0} \geq \left(\partial W^S/\partial N_A \cdot N_B \cdot p' \right)_{\theta=0}. $$

(2.8)

Notice that when the above inequality is strict $\theta^* > 0$ strictly dominates $\theta^* = 0$, and when the equality holds the professor is indifferent between the two strategies. Furthermore, one can readily see that when student A plays fair and student B cheats, the severe professor’s best choice is $\theta^* > 0$ if and only if $(p' \cdot \partial W^S/\partial p + \partial W^S/\partial \theta)_{\theta=0} > (\partial W^S/\partial N_B \cdot N_A \cdot p')_{\theta=0}$.

### 2.1.3 Nash Equilibria

We have already found all the players’ best choices. Now we are able to compute all the several possible Nash equilibria of the baseline game. However, most of those outcomes involve at least one student cheating. In this section we are interested in finding conditions to guarantee existence and uniqueness of what we call the *virtuous equilibrium*, a Nash
equilibrium in which both students play fair. In order to do this, we first establish an important condition:

\[
\left( p \frac{\partial W^S}{\partial p} + \frac{\partial W^S}{\partial \theta} \right)_{\theta=0} \geq \max \left\{ \begin{array}{c} \frac{\partial W^S}{\partial N_A} N_B p' \left\vert_{\theta=0} \right. \\ \frac{\partial W^S}{\partial N_B} N_A p' \left\vert_{\theta=0} \right. \end{array} \right. \right) ,
\]

which can be seen equal to (2.8) when the students are symmetrical - in this case \( \partial W^S \partial N_B N_A \cdot p' = \partial W^S \partial N_A N_B \cdot p' \).

Our main result in this section is given by the next proposition\(^3\).

**Proposition 2.9** The baseline game played by student A, student B, and the professor has the virtuous equilibrium if and only if \( p^* \geq p^{\text{min}} \), the professor is severe and condition (2.9) holds. Furthermore, if the above inequality is strict and condition (2.9) holds with strict inequality, then the equilibrium is unique.

The idea underlying the above result is that the virtuous equilibrium can only happen if professor gives students incentives to behave honestly. This is done by increasing the probability of detecting cheating above the threshold \( p^{\text{min}} \), which in turn requires that the professor make a positive effort, which it is impossible when he is lenient. Thus, the equilibrium with no cheating is only possible when the professor is severe and his marginal disutility from the decreasing grades is relatively low, as established by condition (2.9). The virtuous outcome is the only Nash equilibrium of the game whenever none of the players is indifferent between their strategies, which is guaranteed when \( p^* > p^{\text{min}} \) and condition (2.9) is satisfied with strict inequality.

### 3 Extensions

#### 3.1 Harder punishment

Consider the case when there exists a further punishment for the dishonest student. For example, assume that the institution (school, college or university) has a code of ethical conduct (academic honor code) that establishes punishments such as failing grade in the course, suspension or expulsion. Let \( F > 0 \) be the constant disutility of this punishment, such that now the student’s expected utility when he chooses to cheat is \( U_i (N_j(e^*_j)(1 - p^*), 0) - p^* F \). Assume also that this further punishment does not affect the professor’s utility - there is no cost to implement it, for example.

The best choice analysis here is quite similar to that of section 2.1.1. In fact, when \( \theta^* = 0 \), there is no difference in the equilibria of the above matrices: given the professor’s choice, student A plays fair and student B cheats and; student A cheats and student B

\(^3\)Proofs of propositions 2.9 and 3.4 are in the text, thus they are not shown in the appendix A like the demonstrations of the other results.
plays fair. However, notice that when $\theta^* > 0$ students face a further incentive to play fair, namely the punishment $F$. This new feature does not change the way they make their best choices. For example, student $A$ chooses to play fair if and only if

$$U_A(N_A(e^*_A), e^*_A) \geq U_A(N_B(e^*_B)(1 - p^*), 0) - p^*F.$$  \hspace{1cm} (3.1)

Therefore, given $p$, the equilibria of the second payoff matrix are still the same, but now the value of $p^{\text{min}}$ is different, as the next result states.

**Proposition 3.1** Suppose that if the student is caught cheating, his grade is set to zero and he is punished by losing $F > 0$ of utility. Then there exists a probability of being caught cheating $\hat{p}^{\text{min}} \in (0, 1)$ such that

$$U_i(N_i(e^*_i), e^*_i) = U_i(N_j(e^*_j)(1 - \hat{p}^{\text{min}}), 0) - \hat{p}^{\text{min}}F,$$  \hspace{1cm} (3.2)

for $i, j = A, B$ and $i \neq j$. Moreover, $d\hat{p}^{\text{min}}/dF < 0$, and in particular, $\hat{p}^{\text{min}} < p^{\text{min}}$, where $p^{\text{min}}$ is defined in the proposition 2.8.

In fact, this further punishment may be large enough to make both students choose to play fair regardless the probability of being caught cheating, as the next corollary shows.

**Corollary 3.2** For any given probability of catching students cheating $p \in (0, 1]$, there exists a punishment level $F^{\text{min}} > 0$ such that if $F > F^{\text{min}}$, then playing fair is a strictly dominant strategy for both students.

The above results are similar to those of seminal study of Becker (1968), in particular, if the cost of committing an offense increases, ceteris paribus, potential offenders will be less prone to do it. Further, the punishment may be large enough to make all of them choose not to commit the offense. This may make us conclude that having a fair class, without cheating, is just a matter of choosing the correct level of punishment $F$. However, there is an underlying assumption in our framework that may be contested, namely there is no cost for the professor to implement this further punishment. As the empirical literature reports (e.g. McCabe et al., 2001), a considerable number of professors claim to treat in-class cheating lightly because of the bureaucratic costs associated to all the steps of a process of punishment. Relaxing such an assumption seems to be an promising extension, which would allow to better understand the professor’s behavior.

### 3.2 Heterogeneous students

Let us now relax the assumption of symmetrical students. If students are no longer identical, they choose different levels of optimal effort. Without loss of generality, we assume that $e^*_A > e^*_B$. This can happen when they have similar utility functions, but
Proposition 3.3 Suppose that \( A \) that the existence of such an minimum probability is not guaranteed for student \( A \) between students, we now call this minimum probability \( p \) and plays fair if student \( A \) being caught cheating \( \theta \) model: when professor chooses \( \theta \) is higher than the one from cheating, regardless the professor’s effort and the probability of his disutility of effort as compared to the benefit from the increased grade. We must detail this difference below.

First observe that student \( B \) behaves exactly in the same way he does in the baseline model: when professor chooses \( \theta^* = 0 \), he plays fair if student \( A \) cheats, and cheats if student \( A \) plays fair, and; when professor chooses \( \theta^* > 0 \), he plays fair if student \( A \) cheats, and plays fair if student \( A \) plays fair if and only if \( p^* \geq p^{min} \). Because of the difference between students, we now call this minimum probability \( p_B^{min} \). The next result shows that the existence of such an minimum probability is not guaranteed for student \( A \).

**Proposition 3.3** Suppose that \( N_A(e_A^*) > N_B(e_B^*) \). Then there exists a probability of being caught cheating \( \hat{p}_A^{min} \in (0, 1) \) for student \( A \) such that

\[
U_A(N_A(e_A^*), e_A^*) = U_A(N_B(e_B^*)(1 - \hat{p}_A^{min}), 0),
\]

if and only if \( U_A(N_A(e_A^*), e_A^*) < U_A(N_B(e_B^*), 0) \). Whenever \( \hat{p}_A^{min} \) exists, we have \( \hat{p}_A^{min} < \hat{p}_B^{min} \).

Taking into account that \( \partial U_A / \partial N_A : N_A' > 0 \) and \( \partial U_A / \partial e_A < 0 \), the condition above is equivalent to saying that student \( A \)’s marginal disutility of effort must be high enough to overcome the marginal benefit from the increase in his grade, when his level of effort is zero and his grade is \( N_B(e_B^*) \). Whenever this condition fails, the utility from playing fair is higher than the one from cheating, regardless the professor’s effort and the probability associated to this effort. In this case, when professor chooses \( \theta^* > 0 \), playing fair is a dominant strategy for student \( A \), since we know that he makes the same choice if the another student cheats (recall that \( U_A(N_A(e_A^*), e_A^*) > 0 \)).

The magnitude of the difference between their grades - and therefore between their levels of effort - is an important component of the above analysis. Let us show this by using extreme examples. Suppose initially that \( N_A(e_A^*) \) is fixed and \( N_B(e_B^*) = 0 \). One can readily see that \( U_A(N_A(e_A^*), e_A^*) > U_A(0, 0), \) such that student \( A \)’s best choice is to play fair, and this is completely independent of \( p^* \). However, if \( N_B(e_B^*) = N_A(e_A^*) \), we have seen that \( U_A(N_A(e_A^*), e_A^*) < U_A(N_A(e_A^*), 0) \), which indicates a possibility of cheating, depending on the probability of being caught. The intuition underlying proposition 3.3 is that student \( A \)’s grade may be so higher than the one of his classmate that the risk of cheating is not worth taking, even when there are very low chances of being punished.
Proposition 3.3 also states that when $p_{\min}^A$ exists, it is lower than $p_{\min}^B$. Once again, since $N_A(e^*_A) > N_B(e^*_B)$, student A’s gain by making a positive effort is higher than the one of the another student, as a result he is “more prone” to play fair than B. This is reflected in the minimum probability that induces the student to behave honestly. Thus, professor’s best choice $\theta^* > 0$ can now result in several possibilities: $p^* > p_{\min}^B > p_{\min}^A$, $p^* = p_{\min}^B > p_{\min}^A$, $p^* > p_{\min}^B = p_{\min}^A$, $p^* > p_{\min}^B > p_{\min}^A$, $p_{\min}^B > p^* > p_{\min}^A$, and $p_{\min}^B > p^* > p_{\min}^A$. As the FOC of professor’s problem (2.2) shows, the chosen case will depend on his marginal utilities, mainly his disutility of effort. For example, if condition (2.2) holds when $-\partial W^S/\partial \theta$ is very large, then the concavity of $W$ implies that $\theta^*$ will be very low, such that we may have the fifth case.

The potential nonexistence of $p_{\min}^A$ creates a multiplicity of Nash equilibria in the cheating game with heterogeneous students, a number even larger than the one of the baseline model. Due to this, we constrain our main analysis only for the case in which $p_{\min}^A$ exists. When $U_A(N_A(e^*_A), e^*_A) \geq U_A(N_B(e^*_B), 0)$, student A has a weakly dominant strategy, namely playing fair, such that it suffices to study the best choices of the other two players, and the results are quite similar to those of proposition 2.9.

Proposition 3.4 Suppose that $e^*_A > e^*_B$. Then the game played by heterogeneous students $A$ and $B$ and the professor has a virtuous equilibrium if and only if $p^* \geq p_{\min}^B$, the professor is severe and condition (2.9) holds. Furthermore, if the above inequality is strict and condition (2.9) holds with strict inequality, then the equilibrium is unique.

The intuition of the above result is quite similar to one of the proposition 2.9, except by the difference between the minimum probabilities of being caught cheating necessary to make students $A$ and $B$ to play fair. Once again, a virtuous equilibrium requires that the professor exert a level of effort high enough to the probability associated be higher than that minimum level. As the higher the grade the lower the $p_{\min}$, a class composed by high effort students - because of their low disutility of effort, for example - makes the professor’s task of maintaining fairness easier.

A possibility that is not exploited in this paper is the presence of more than two students - identical or not - in the classroom. This modification would make each student has more potential “victims”, but at the same time there would be more potential “offenders” as well. Issues such as how to set students’ seating position in order to minimize the chance of cheating would be possible to be studied when there are more players in the game. As we have seen in this section, if these students were heterogeneous, a plenty of possible outcomes would emerge and the differential between students’ grades would have an important role.
4 Concluding remarks

The novelty and main contribution of this paper is to highlight that cheating may be seen as a strategic choice, which involves cost-benefit analysis. In fact, our framework provides the microeconomic foundations of both student’s choice of cheating or not and professor’s choice of trying to catch dishonest students. By applying game theory’s tools, we are able to better understand the determinants of academic dishonesty found by the literature, in a similar way the theoretical model of Becker (1968) have done with Economics of Crime. Our findings also provide further policy implications for cheating control in classroom. In particular, we emphasize the importance of professors being well-motivated (with low disutility of effort) and worried about fairness in classroom.

Our model is the first step towards a rigorous treatment of the strategic relationships underlying the students’ choice of cheating or not. Therefore, there are several directions in which it can be extended. One that we believe to be promising is relaxing the assumption of complete information. A student may be unsure about the true probability of being caught cheating, which can be modeled by assuming that he does not know the professor’s level of effort (professor’s type), only has a belief about it. He may also be unsure about the other student’s true grade - or his true effort level, his type ultimately -, which can be modeled in the same way. Such an extension would allow the study of issues such as signaling, which in turn allows us to understand how professor can use the daily contact with students to prevent cheating.

References


### A Omitted proofs

#### A.1 Proposition 2.2

Observe that

\[
\lim_{e_i \to +\infty} \frac{dU_i}{de_i} = \lim_{e_i \to +\infty} \left( \frac{\partial U_i}{\partial N_i} N_i' \right) + \lim_{e_i \to +\infty} \frac{\partial U_i}{\partial e_i}.
\]  

(A.1)

Thus, by conditions (iii) and assumption 2.1, \(\lim_{e_i \to +\infty} dU_i/de_i = \lim_{e_i \to +\infty} \partial U_i/\partial e_i = -\infty\). Recall that \(U_i\) is \(C^2\), such that \(dU_i/de_i\) is continuous. Moreover, by condition (i), \(dU_i/de_i(0,0) > 0\). Thus, we can invoke the intermediate value theorem and conclude that there exists an interior point \(e_i^*\) that satisfies (2.1).

We must now show that that \(e_i^*\) is an unique global maximizer of the student’s optimization problem. For, note that

\[
\frac{d^2U_i}{de_i^2} = \frac{\partial^2 U_i}{\partial N_i^2} (N_i')^2 + \frac{\partial U_i}{\partial N_i} N_i'' + 2 \frac{\partial^2 U_i}{\partial N_i \partial e_i} N_i' + \frac{\partial^2 U_i}{\partial e_i^2} < 0,
\]  

(A.2)
for all $e_i$, because of condition (ii), assumption 2.1 and $\partial U_i/\partial N_i > 0$. This implies that $U_i$ is strictly concave in $e_i$. Therefore, the first order condition is sufficient to guarantee that $e^*_i$ is an unique global maximizer. ■

A.2 Proposition 2.3

To prove the first claim of the proposition, suppose that student $A$’s best choice $e^*_A$ satisfies (2.1). Suppose also that $\partial U_A/\partial N_A = \partial U_B/\partial N_B$ and $N'_A = N'_B$, but $\partial U_A/\partial e_A > \partial U_B/\partial e_B$. This implies that at $e_B = e^*_A$

$$\frac{dU_B}{de_B} = \frac{\partial U_B}{\partial N_B} N'_B + \frac{\partial U_B}{\partial e_B} < 0. \quad (A.3)$$

Thus, given that $d^2U_i/d e^2_i < 0$, we must have $e^*_B < e^*_A$.

Second and third claims can be proved by using the same reasoning. For, suppose that $\partial U_A/\partial N_A > \partial U_B/\partial N_B$, $N'_A = N'_B$, and $\partial U_A/\partial e_A = \partial U_B/\partial e_B$. Now observe that at $e_B = e^*_A$ we once again have (A.3), such that we can conclude that $e^*_B < e^*_A$. It is straightforward to see that the same inequality is found when we suppose that students have the same marginal utilities, but different returns of the effort on grades. ■

A.3 Proposition 2.6

Item (i) of assumption 2.4 implies $dW^S/d\theta$ is strictly decreasing in $\theta$. Moreover, by the definition of severe professor, $dW^S/d\theta > 0$ when $\theta^* = 0$. From item (ii) of assumption 2.4 we also have \(\lim_{\theta \to +\infty} dW^S/d\theta = -\infty < 0\). Finally, given that $W$ is a $C^2$ function, its derivative is continuous. Thus, the intermediate value theorem applies and there exists an interior point $\theta^*$ that satisfies (2.2). This point is an unique global maximizer because $W$ is strictly concave in $\theta$. ■

A.4 Proposition 2.7

We employ the same reasoning of the proposition’s 2.3 proof. We prove the proposition for the case when student $A$ cheats and student $B$ plays fair. Proofs for the remaining cases are straightforward and very similar to this one. Suppose that $\theta^*$ satisfies the severe professor’s FOC (2.2). For the item (i), assume that all the derivatives of his utility function are fixed except his marginal disutility of the effort, which now is $\partial W^S/\partial \theta < \partial W^S/\partial \theta$. This implies that at $\theta = \theta^*$

$$p' \left( \frac{\partial W^S}{\partial p} - \frac{\partial W^S}{\partial N_A} N_B \right) + \frac{\partial W^S}{\partial \theta} < 0. \quad (A.4)$$
Thus, given that $d^2W^S/d\theta^2 < 0$, we must have $\hat{\theta} < \theta^*$, where $\hat{\theta}$ solves $p' \left( \partial W^S \over \partial p - \partial W^S \over \partial N_A N_B \right) + \partial W^S \over \partial \theta = 0$.

For item (ii), assume that the only derivative that is not fixed is $\partial W^S/\partial N_A$, which now is $\partial W^S/\partial N_A > \partial W^S/\partial N_B$. One can see that at $\theta = \theta^*$ we once again have $dW^S/d\theta < 0$, which implies $\hat{\theta} < \theta^*$. We can repeat the procedure for the other two items and find that $dW^S/d\theta > 0$ when $\theta = \theta^*$. Therefore, in those cases $\hat{\theta} > \theta^*$. ■

A.5 Proposition 2.8

First, define a function $f : [0, 1] \to \mathbb{R}$, given by $f(p) := U_i \left( N_i(e^*_i), e^*_i \right) - U_i \left( N_j(e^*_j)(1 - p), 0 \right)$. Then, observe that $f$ is continuous, because so is $U_i$, and

$$f'(p) = \frac{\partial U_i}{\partial N_i} p N_i > 0, \quad (A.5)$$

that is, $f(p)$ is strictly increasing for all $p \in [0, 1]$.

Now we can compute

$$f(0) = U_i \left( N_i(e^*_i), e^*_i \right) - U_i \left( N_j(e^*_j), 0 \right)$$

$$= U_i \left( N_i(e^*_i), e^*_i \right) - U_i \left( N_i(e^*_i), 0 \right) < 0 \quad (A.6)$$

$$f(1) = U_i \left( N_i(e^*_i), e^*_i \right) - U_i \left( 0, 0 \right)$$

$$= U_i \left( N_i(e^*_i), e^*_i \right) > 0, \quad (A.7)$$

where we use the symmetry of the students and the fact that $U_i(0, 0) = 0$. Therefore, given that $p \in [0, 1]$, the continuity of $f$ and $f'(p) > 0$, then there exists $p^{\text{min}} \in (0, 1)$ such that $f(p^{\text{min}}) = 0$, which is what had to be proven. ■

A.6 Proposition 3.1

The proof is similar to that of proposition 2.8. First, define a function $g : [0, 1] \to \mathbb{R}$, given by $g(p) = U_i \left( N_i(e^*_i), e^*_i \right) - U_i \left( N_j(e^*_j)(1 - p), 0 \right) + p F$ and observe that $g(p) = f(p) + p F$, where $f$ is defined in proposition 2.8. In addition, $g(p)$ is strictly increasing because $g'(p) = f'(p) + F > 0$ for all $p \in [0, 1]$. By computing $g(0) = f(0) < 0$ and $g(1) = f(1) + F > 0$, and recalling that $g(p)$ is continuous, because so is $f(p)$, we can once more invoke the intermediate value theorem to conclude that there exists a $p^{\text{min}} \in (0, 1)$ such that $g(p^{\text{min}}) = 0$.

We must now show that $p^{\text{min}}$ is decreasing in $F$. For the particular case $p = p^{\text{min}}$, we have $g(p^{\text{min}}) = f(p^{\text{min}}) + p^{\text{min}} F = p^{\text{min}} F > 0$, where $p^{\text{min}}$ is also defined in proposition 2.8. Given that $g'(p) > 0$ for all $p$, it must be the case that $p^{\text{min}} < p^{\text{min}}$. For the general case,
one can see that
\[
\frac{d\hat{p}^{\min}}{dF} = -\frac{\partial g/\partial F}{\partial g/\partial \hat{p}^{\min}} = -\frac{\hat{p}^{\min}}{F} < 0,
\] (A.8)
where we used the implicit function theorem. ■

A.7 Corollary 3.2

Let \( \bar{p} \in (0, 1] \) be a given constant. Now we consider the function \( h : \mathbb{R}^* \to \mathbb{R} \), given by
\[
h(F) = U_i(N_i(e_i^*), e_i^*) - U_i(N_j(e_j^*)(1 - \bar{p}), 0) + \bar{p}F.
\]
Define
\[
F^{\min} = \frac{\left[U_i(N_j(e_j^*)(1 - \bar{p}), 0) - U_i(N_i(e_i^*), e_i^*)\right]}{\bar{p}}.
\] (A.9)

Observe that if \( F > F^{\min} \), then \( h(F) > 0 \), that is, playing fair is the best choice when the another student plays fair as well. Given that student \( i \) also chooses to play when student \( j \) cheats, the strategy is dominant when the above condition holds.

Now we have to consider three cases. First, if \( \bar{p} > p^{\min} \), then the numerator of the above expression is negative, and so is \( F^{\min} \). Therefore, in this case any punishment \( F \geq 0 \) guarantees that playing fair is a dominant strategic for both students. Second, if \( \bar{p} = p^{\min} \), then \( F^{\min} = 0 \), such that any positive punishment is sufficient for the result. Finally, if \( \bar{p} < p^{\min} \), then \( F^{\min} > 0 \) and we need that \( F > F^{\min} \). ■

A.8 Proposition 3.3

First, suppose that \( U_A (N_A(e_A^*), e_A^*) < U_A (N_B(e_B^*), 0) \) and consider again function \( f(p) := U_i(N_i(e_i^*), e_i^*) - U_i(N_j(e_j^*)(1 - p), 0) \) with \( i = A, j = B \) defined in the proof of proposition 2.8. Then, we have \( f(0) < 0 \) and \( f(1) > 0 \). Given that \( f \) is \( C^2 \) and \( f' > 0 \) for all \( p \), there exists a \( \hat{p}_A^{\min} \in (0, 1) \) such that \( f(\hat{p}_A^{\min}) = 0 \).

Now, suppose that there exists a \( \hat{p}_A^{\min} \in (0, 1) \) such that \( f(\hat{p}_A^{\min}) = 0 \). Then, for any \( \epsilon > 0 \) small enough, \( f(\hat{p}_A^{\min} + \epsilon) > 0 \) and \( f(\hat{p}_A^{\min} - \epsilon) < 0 \), since that \( f' > 0 \) for all \( p \). In particular, when \( \epsilon = \hat{p}_A^{\min} \), we have \( f(\hat{p}_A^{\min} - \epsilon) = f(0) < 0 \), which implies \( U_A (N_A(e_A^*), e_A^*) < U_A (N_B(e_B^*), 0) \).

The final step of the proof is to demonstrate that \( \hat{p}_A^{\min} < \hat{p}_B^{\min} \). This can be shown by calculating the following derivative:
\[
\frac{d\hat{p}^{\min}}{dN_i(e_i^*)} = -\frac{\partial f/\partial N_i(e_i^*)}{\partial f/\partial \hat{p}^{\min}} = -\frac{1}{N_j(e_j^*)(1 - p)} < 0,
\] (A.10)
that is, the minimum probability of student \( i \) decreases when his own grade increases, ceteris paribus. Therefore, if when the students are identical we have \( \hat{p}_A^{\min} = \hat{p}_B^{\min} \), now the one with higher grade must have a minimum probability lower than his classmate. ■