

# ANÁLISIS DE LA AUTOLESIÓN COMO HERRAMIENTA DE FRAUDE

Enrique Martín Luccioni<sup>1</sup>

Julio, 2015.

## Resumen

Este trabajo presenta una contribución al análisis de la autolesión para fines sociales. Mientras que la literatura actual de agentes racionales se concentra en modelos de señales para explicar dicho comportamiento en prisión, aquí se desarrollará un modelo de fraude donde la autolesión disminuye la probabilidad efectiva de ser descubierto simulando un estado accidental de la naturaleza. Se demostrará que el fraude puede ser una alternativa viable a un juego clásico de pedido de ayuda en prisión aun cuando exista un equilibrio separador en este último. Se demostrará además las condiciones necesarias para la existencia de un equilibrio en ambos juegos.

*Clasificación JEL: D81, D82, Z13*

*Palabras clave: Autolesión, fraude, prisión, pedido de ayuda, screening.*

---

<sup>1</sup> Autor: Enrique Martín Luccioni. Correo electrónico: [luccioni.enrique@inta.gob.ar](mailto:luccioni.enrique@inta.gob.ar). Agradezco a Daniel Lema, Jorge Streb, Julio Elias y Osvaldo Meloni por sus comentarios y sugerencias. También a Karina Casellas, Nicolas Gatti, Ignacio Pace y el resto del equipo del Instituto de Economía del INTA por sus aportes y sugerencias.

## **Introducción**

Se cree que todos los animales tienen un impulso natural por sobrevivir (Dawkins 2006), resulta difícil entonces comprender las razones que llevan a los humanos a destruir su propio cuerpo. A primera vista la autolesión sin fines suicidas aparenta ser fruto de un desorden mental, sin embargo este no es necesariamente el caso, por ejemplo este trabajo explicará el comportamiento de un prisionero que decide simular un accidente para ser trasladado y evitar ser atacado.

Gran parte de la literatura sociológica y biológica estudia la autolesión desde la óptica de agentes racionales. Esto presenta un desafío mayor, ya que si se supone consistencia y monotonía en las preferencias, ¿Cómo es posible entonces que las personas decidan lastimarse? El modelo de señal costosa en un entorno donde las palabras no implican un compromiso creíble es compartido por varios autores (Hagen et al. 2008, Kaminski 2004, Gambetta 2009, Griller 2013).

A diferencia del enfoque de señales este trabajo busca modelar el comportamiento de un individuo que simula un accidente para obtener un beneficio o evitar una pérdida. Como estrategia para disminuir la probabilidad de ser descubierto el individuo cuenta con la opción de autolesionarse. Recurriendo a la teoría del fraude económico, la autolesión será modelada como una herramienta para engañar un proceso de auditoría. El fraude no requiere un intercambio voluntario ya que solamente una de las partes podría beneficiarse del mismo.

Para diferenciar los enfoques se demostrará que la autolesión como herramienta de engaño puede servir para los mismos fines que un juego de *Screening* donde la parte no informada exige una prueba costosa y obtiene un beneficio si se transmite la verdad. A modo de ejemplo, un niño que desea llamar la atención de su madre la cual no reacciona

ante su llanto puede decidir autolesionarse para demostrarle su angustia o puede simular un accidente lo que le permite conseguir la atención que buscaba.

El enfoque de autolesión desde la teoría del fraude es mencionado tangencialmente por varios autores pero nunca modelado, por ejemplo Kaminiski (2004) explica la utilidad de la simulación en prisión y cita comunicación personal con un prisionero político que utilizó simulación y autolesión para ser transferido. Griller (2013) menciona la posibilidad de aparentar heridas más severas de lo que realmente son, lo que podría ser modelado con el enfoque de falsificación de Crocker y Morgan (1997) donde un individuo magnifica el daño de un accidente.

El modelo presentado puede explicar desde el comportamiento de desertores en el campo de batalla con heridas autogeneradas hasta un niño que decide faltar a un examen escolar o cualquier comportamiento donde exista una auditoría que puede ser evitada destruyendo capital propio. En la Parte I se introduce el problema de la autolesión desde un enfoque de agentes racionales. En la parte II III y IV se formaliza la autolesión desde la óptica de *screening* y de fraude. En la parte V se comparan los resultados obtenidos para cada modelo. En la parte IV se discute la generalidad del enfoque.

### **Parte I. Autolesión como proceso de decisión racional**

La autolesión no suicida se define como un daño directo y deliberado del cuerpo humano con ausencia de intenciones suicidas y con propósitos no aceptados socialmente<sup>2</sup> (Nock 2009, Klonsky et al 2011). Nock & Prinstein (2004) presentan un modelo funcional donde definen dos categorías dicotómicas, en primer lugar las contingencias para autolesionarse pueden ser intrapersonales o sociales, en segundo

---

<sup>2</sup> Los desórdenes alimenticios no entran dentro de este grupo porque el daño es indirecto y no inmediato. El masoquismo tampoco ya que normalmente necesita la participación activa de otra persona. Finalmente tatuajes o piercings no entran dentro de la categoría ya que actualmente son socialmente aceptados.

lugar el reforzamiento puede ser positivo (seguido por un estímulo favorable) o negativo (seguido por la eliminación de un estímulo desfavorable).

Nock (2008) presenta una subdivisión adicional dentro de la categoría social al considerar la autolesión como señal de angustia o pedido de auxilio y como señal de fuerza o buen estado físico. Literatura más reciente (Bentley et al. 2014) admite que las funciones sociales no se encuentran profundamente investigadas y aclara que es necesario estudiar los procesos que llevan a los individuos a autolesionarse para fines interpersonales. En este sentido, el presente trabajo busca demostrar que una potencial alternativa al modelo de señales es considerar la autolesión como una herramienta para simular un accidente. La tabla I presenta un resumen ejemplificador de las funciones sociales incluyendo los modelos de señales anteriormente citados y posibles usos de la autolesión como engaño.

Tabla I – Funciones sociales de la autolesión

	funciones sociales	
	reforzamiento positivo	reforzamiento negativo
Señal de angustia	Llamar la atención de otro demostrando que está dispuesto a lesionarse	Evitar una tarea demandante (ir al colegio) demostrando que está dispuesto a lesionarse
Señal de coraje	Impresionar a otros, conseguir pareja	"cicatrices de combate" para disuadir atacantes
Engaño	Llamar la atención de otro simulando un accidente	Evitar una tarea demandante simulando un accidente

Fuente: Elaboración propia a partir de Nock (2008)

Bajo la óptica de agentes racionales Kaminski (2004) realiza un estudio de la autolesión a partir de su experiencia personal como prisionero político. Considera dos usos estratégicos para la misma, por un lado la autolesión como señal y por otro la simulación de un accidente, sin embargo solo modela el primero de ellos. Además distingue entre señales simbólicas donde maximizan ambos jugadores y la autolesión

pragmática donde el carcelero actúa automáticamente ante un evento, por ejemplo un guardia que se encuentra ante un intento de suicidio debe llamar por reglamento a una ambulancia, por lo que su rol es puramente técnico y no debe optimizar una decisión.

Siguiendo con la bibliografía sociológica, Gambetta (2009) racionaliza el comportamiento de prisioneros que utilizan la autolesión como una señal estratégica y presenta las condiciones necesarias para la existencia de autolesión como señal costosa de coraje. Griller (2013) presenta una formalización de los juegos anteriormente citados y un refinamiento de los requisitos para su existencia. Bajo el enfoque de biología evolutiva Hagen et. Al (2008) presenta una sofisticación al modelo clásico de señales suponiendo un modelo de negociación de múltiples periodos. Bajo esta hipótesis si los costos de elegir e intercambiar pareja son muy altos entonces la autolesión puede ser un mecanismo eficiente para presionar a la pareja para obtener concesiones.

El trabajo de medicina forense de Heide y Kleiber (2006) agrupa la autolesión en tres conjuntos, psicológicos, judiciales y materiales. Los dos últimos pueden ser fácilmente modelados con el enfoque de fraude que se da en este trabajo, por ejemplo dentro de los motivos judiciales se encuentra la simulación de un crimen penal y la defensa contra reproches (justificar inasistencias). Los motivos materiales son por ejemplo fraudes en seguros, la evasión del servicio militar o la obtención de mejores condiciones de vida en prisión al ser transferido a un hospital.

Es inevitable realizar abstracciones para comprender el comportamiento racional de los autolesionados<sup>3</sup>. Las partes del cuerpo humano son un capital. Sin embargo no posible acceder a un mercado convencional donde voluntariamente se intercambia dinero por capital (por ejemplo al comprar un automóvil), principalmente porque la tecnología

---

<sup>3</sup> Estas abstracciones pueden ser catalogadas de crueles o desalmadas, sin embargo su objetivo es demostrar que las decisiones racionales de estos individuos no es distinta a la del resto de las personas, incluidas por supuesto los lectores de este artículo.

actual no lo permite y aun si lo permitiese los individuos rechazarían participar del mismo al resultarles repulsivo (Roth 2007).

Sin embargo ciertas partes del cuerpo humano si pueden ser intercambiadas voluntariamente, por ejemplo es conocida la existencia de un mercado informal de venta de riñones en donantes vivos, Becker y Elias (2007) analizan los efectos económicos de dicha posibilidad.

## **Parte II. Juego de *Screening*: pedido de auxilio costoso.**

A continuación se desarrolla un modelo de autolesión con información asimétrica donde el primer movimiento es realizado por el jugador no informado el cual exige una prueba, conocido en la literatura como juegos de *Screening* (Rasmusen 2007).

Para simplificar la interpretación se utilizará una versión modificada de un juego señales en la literatura sociológica (Gambetta 2009, Kaminski 2004 y Griller 2013) donde un preso está a punto de ser atacado, los carceleros no pueden observar de forma directa si el ataque ocurrirá, por ende para llamar la atención de los guardias el preso decide autolesionarse.

Los carceleros producen diversos bienes dentro de prisión, uno de ellos es la seguridad de los presidiarios que es recompensada según su efectividad para impedir ataques. Dado que los carceleros no pueden observar de forma directa el daño potencial que pueden sufrir los prisioneros, necesitan recurrir a información adicional que les permitan distinguir amenazas reales, la autolesión aunque cruel representa un mecanismo idóneo para esta función ya que solo los prisioneros que se encuentren en verdadero peligro van a recurrir a ella. En términos de Adam Smith, no es la benevolencia del carcelero lo que lo incentiva a ayudar a los presidiarios si no el pago que espera recibir en compensación.

El carcelero no puede confiar en la honestidad de los presidiarios para revelar su verdadero estado de la naturaleza, por ejemplo un prisionero deshonesto que no va a sufrir ningún ataque puede desear ser trasladado a un lugar mejor. Además, lo que se define como una situación peligrosa está regido por lo que la administración penitenciaria considere nocivo para sus intereses y por ende es posible que aun cuando un preso esté por sufrir un ataque real este no sea lo suficientemente importante para llamar la atención del carcelero. Es por esto que la falta de información y la diferencia entre los intereses entre presos y carceleros socava la transmisión de mensajes honestos.

Un juego de *screening* es un juego de información incompleta con dos o más jugadores, uno de ellos informado de una cualidad no observable directamente por el otro jugador. La principal ventaja de un juego de *screening* sobre uno de señales clásicos es que el jugador no informado mueve primero, al no existir creencias en el equilibrio se eliminan los equilibrios de tipo agrupador (Rasmusen 2007). Para modelar la asimetría de información se supondrá que existe un diferencial en los beneficios entre aquellos que sufren un gran peligro y el resto de los presidiarios<sup>4</sup>. A continuación se esquematizaran las características claves del juego siguiendo el enfoque de Gambetta (2009b). Para ello se definirán los jugadores, la propiedad oculta que genera la asimetría de información y la ganancia de los participantes, por último se definirá el orden del juego.

Existe un prisionero que sabe si el peligro que corre llamará la atención de dos carceleros que serán los jugadores no informados<sup>5</sup>. Además existen jugadores indirectos, la administración carcelaria que paga al carcelero en la medida que puede evitar un ataque y los potenciales agresores del prisionero. Existe además una propiedad oculta, los carceleros no pueden observar el peligro potencial que corre el presidiario.

---

<sup>4</sup> Una comparación de modelos de señales con diferenciales de costos y diferencial de beneficios puede encontrarse en Gambetta (2009b).

<sup>5</sup> De forma análoga al supuesto de dos firmas en el modelo de Spence (1974) se supondrá competencia a la Bertrand entre los dos carceleros para eliminar la renta monopólica

Para simplificar se supondrá que existen dos niveles de ataque con una retribución asociada para el carcelero sólo en el caso de que el ataque sea grande.

La ganancia para el prisionero si se transmite la verdad será evitar las agresiones al ser trasladado a otro sector de la cárcel por lo que aún aquellos prisioneros que sufren un ataque menor tendrán incentivos a llamar la atención de los guardias<sup>6</sup>. Si un prisionero con peligro bajo es trasladado entonces el carcelero no tendrá ninguna ganancia.

El orden del juego es el siguiente. Primero la naturaleza decide el nivel de ataque, en segundo lugar ambos carceleros hacen público el nivel de esfuerzo que realizará dependiendo de la autolesión observada. En tercer lugar el prisionero decide el nivel de autolesión que se infringirá. Por último el prisionero puede elegir entre ser ayudado (si cumple los requisitos de autolesión en el segundo paso) o rechaza la ayuda.

La riqueza inicial de los prisioneros es  $W$ , al existir dos niveles de ataque futuro la naturaleza decide una pérdida  $l_i \in \{l_b, l_a\}$  con  $0 < l_b < l_a$  y con una distribución de tipos  $\sigma$  si el daño es alto y  $(1 - \sigma)$  si es bajo, esta pérdida tiene asociada una retribución monetaria para el carcelero si la impide dada por  $a_{(l_b)} = 0$  y  $a_{(l_a)} = a^* > 0$ . Los carceleros tienen asociada una función de costo  $e_{(D)}$ . Los prisioneros eligen un nivel de autolesión  $D \in [0, \infty)$  y tienen una pérdida  $m_{(e)}$  que depende del nivel esfuerzo de los carceleros, si los carceleros no ayudan entonces  $m_{(e=0)} = l_i$ <sup>7</sup>, mientras que si deciden ayudar  $m_{(e>0)} = 0$ . Se supondrá funciones de utilidad lineales con subíndices que reflejan si el prisioneros sufrirá un ataque alto o bajo.

---

<sup>6</sup> Se supondrá que no existen beneficios adicionales por ser trasladado, para un análisis detallado de los motivos para trasladarse ver Kaminski (2004)

<sup>7</sup> Esta condición genera la asimetría de beneficios que garantizará un equilibrio separador

$$U_{prisionero} = \begin{cases} W - m_i(e) - D_i & \text{si acepta ayuda} \\ W - l_i & \text{si no acepta ayuda} \end{cases} \quad (1)$$

$$U_{carcelero} = \begin{cases} a(l_i) - e(D_i) \\ 0 \end{cases} \quad (2)$$

Lo cual permite encontrar un equilibrio separador

$$\text{Equilibrio Screening} \begin{cases} D(l_{bajo}) = 0, D(l_{alto}) = D^* \\ m = \begin{cases} l_i & \text{si } D < D^* \\ 0 & \text{si } D \geq D^* \end{cases} \quad e = \begin{cases} 0 & \text{si } D < D^* \\ e^* & \text{si } D \geq D^* \end{cases} \end{cases}$$

Que para su existencia deben cumplirse una serie de requisitos, en primer lugar las restricciones de participación de los carceleros que se obtienen a partir de (2) que implican que los carceleros prefieren ayudar a su utilidad de reserva y que estos ofrecerán ayuda sin incurrir en pérdidas.

$$e(0) \leq a(l_b) = 0 \quad (3)$$

$$e(D^*) = e^* \leq a(l_a) = a^* \quad (4)$$

Donde  $D^*$  es el nivel separador de autolesión que se busca. La competencia de Bertrand entre carceleros vuelve las inequidades en equidades por lo que no tendrán beneficios. Siguiendo el orden del juego, los carceleros anuncian que si observa un nivel de daño  $D^*$  se comprometerán a esforzarse  $e^*$  para evitar el ataque, esto garantizará que la pérdida  $m_{(e^*)}$  sea nula.

En segundo lugar las restricciones de autoselección de los prisioneros también llamadas compatibilidad de incentivos induce a los diferentes tipos prisioneros a elegir diferentes contratos. Los de peligro bajo no son atraído al contrato de los de peligro alto y viceversa. Estas condiciones permitirán encontrar el nivel de daño  $D^*$  que indicará el nivel de esfuerzo que realizarán los carceleros. En primer lugar se derivará la condición para los prisioneros con pérdida baja a partir de (1)

$$U_b(D = 0) \geq U_b(D = D^*) \rightarrow W - l_b \geq W - D^* \quad (5)$$

Esta condición se cumplirá con igualdad cuando  $D^* = l_b$  lo que determina el nivel de daño óptimo para encontrar el equilibrio separador. En segundo lugar se deriva la condición para los individuos con pérdida alta

$$U_a(D = 0) \leq U_a(D = D^*) \rightarrow W - l_a \leq W - D^* \quad (6)$$

Para  $D^* = l_b$  la condición es verdadera y será verdadera para valores  $D^* \leq l_a$ , sin embargo el contrato más atractivo será aquel con  $D^* = l_b$  ya que Pareto domina a todos los demás, al existir competencia entre los carceleros se garantiza el contrato más atractivo que cumpla con las restricciones de autoselección y participación

Por último es necesario que se cumpla la restricción de equilibrio agrupador (7) la cual garantiza la existencia de un equilibrio separador. El nivel de esfuerzo esperado que hace cero el beneficio esperado de los carceleros será  $\sigma e^* = \sigma a^* + (1 - \sigma)a_{(l_b)}$  ya que  $a_{(l_b)} = 0$ , lo que determinará la utilidad esperada para los prisioneros con peligro alto en un equilibrio de tipo agrupador. Esta utilidad puede observarse en el lado derecho de la restricción (7).

$$U_a(D = D^*) = W - l_b \geq U_a(\text{agrupador}) = \sigma * (W) + (1 - \sigma) * (W - l_a) \quad (7)$$

Para simplificar el análisis se supondrá arbitrariamente el cumplimiento de (7). Si existe un equilibrio separador entonces la ganancia para un prisionero que esta por sufrir un ataque alto vendrá dada por (8) y puede obtenerse al reemplazar el daño  $D^* = l_b$  en la función de utilidad y restar la utilidad si no acepta ayuda.

$$U_a(D = D^*) - U_a(D = 0) = l_a - D_a^* = l_a - l_b \quad (8)$$

La intuición detrás de este resultado es simple, si el prisionero no acepta ser ayudado tendrá una pérdida  $W - l_a$  mientras que si acepta tendrá una utilidad  $W - l_b$  por lo

tanto la ganancia por autolesionarse vendrá dada por la diferencia entre las pérdidas de los dos tipos de ataque. Es importante notar que existe autolesión en el equilibrio, este resultado será importante para comparar con el modelo de fraude.

#### **Parte IV. Modelo con manipulación de auditoría: Simulación y Autolesión.**

Supongamos ahora que el carcelero no recibe retribución por ayudar a un prisionero antes de un ataque, esto podría ocurrir si el sueldo del carcelero es fijo y no depende ni de la cantidad ni calidad de la ayuda que brinda. En términos del modelo ya presentado esto implica que  $a(l_b) = a(l_a) = 0$ . En dicho caso no se cumplen las restricciones de participación de los carceleros dada por la ecuación (3) y (4) lo que impide la existencia de un equilibrio separador.

Sin embargo se supondrá que los carceleros están obligados por contrato a atender a los prisioneros luego de que sufrieron un ataque o accidente (por ejemplo la legislación carcelaria de Alemania lo obliga Dahle et al 2005), por ende los prisioneros tendrán incentivos a simular un estado accidental de la naturaleza para evitar un peligro real inminente, en la medida que la simulación sea exitosa podrán ser trasladados<sup>8</sup>.

Se supondrá además que los carceleros castigarán a los prisioneros que descubran realizando fraude, de este modo aunque estén obligados a atender los accidentes verídicos, el hecho de que esto no les reporte un beneficio los incentiva a castigar a aquellos que desean aprovecharse de esta situación.

---

<sup>8</sup> Kaminski (2004) menciona el ahorcamiento como un mecanismo usual, un compañero desata al prisionero justo antes de morir, según el modelo presentado este comportamiento puede ser totalmente racional si por ejemplo, evita el asesinato del prisionero.

El modelo está basado en el modelo de fraude y selección adversa de Picard (1996) y Picard (2013) para el mercado de seguro<sup>9</sup>. De forma resumida, se busca modelar la estrategia óptima de auditoría por parte de los carceleros en una prisión donde los prisioneros pueden simular un accidente, para evitar dicho comportamiento los carceleros podrán comprometerse a una política de auditoría con castigo. Se presenta una modificación al modelo original para incluir la posibilidad de que los prisioneros puedan autolesionarse para disminuir la probabilidad de ser detectados.

El presente trabajo al suponer que la autolesión afecta la probabilidad de ser detectado se aparta de los modelos tradicionales de fraude. Bond y Crocker (1997) suponen que la manipulación afecta directamente los costos de detección y Crocker y Morgan (1997) suponen que el asegurado sobredimensiona la lesión. Bouergeon y Picard (2000) desarrollan un modelo de incendios deliberados donde el asegurado tiene información privada sobre el valor real de su propiedad y es esta característica la que motiva la destrucción de sus propios activos, sin embargo el daño auto provocado no afecta la probabilidad de ser detectado.

Se asume al igual que antes un ataque futuro que reporta una pérdida  $l_i \in \{l_b, l_a\}$  con una distribución de tipos  $\sigma$  si el ataque es alto y  $(1 - \sigma)$  si es bajo. Además ocurrirá un accidente con pérdida  $A$  con probabilidad  $\delta \in (0,1)$ . Por su parte los carceleros auditan y detectan fraude exitosamente con probabilidad  $p \in [0,1]$  a un costo  $c$ . Los carceleros deben realizar un esfuerzo  $e$  por atender prisioneros accidentados. Se asume un castigo exógeno  $B$  que deben pagar los prisioneros que son detectados simulando un accidente. Los prisioneros simularán un accidente con una probabilidad endógenamente determinada de  $\alpha_i \in [0,1]$  donde el subíndice refleja si el prisionero sufrirá un ataque

---

<sup>9</sup> En Picard (2013) existe una cobertura que incentiva el fraude y asimetría entre honestos y deshonestos, aquí se supondrá una evitación de la pérdida futura y asimetría en el ataque.

alto o bajo. Se supondrá que no incurrirán en ningún costo al simular. Además, si un prisionero sufrió un accidente es trasladado por lo que no será atacado y no incurrirá en un fraude.

Los prisioneros se infligen un daño  $D_i \in [0, \infty)$  que les producirá una pérdida efectiva de riqueza. A su vez,  $\omega_{(D,A)} \in [0,1]$  es una tecnología que permite disminuir la probabilidad de ser detectado. Se supone  $\omega_{(D=0)} = 1$ ,  $\omega_{(D \geq A)} = 0$  y que  $\frac{\partial \omega}{\partial D} < 0$ . De este modo la probabilidad de auditoría  $p$  se diferenciará de la probabilidad efectiva de descubrir un deshonesto  $p\omega_{(D)}$ , a medida que  $D$  aumenta la probabilidad efectiva de ser descubierto disminuye para un dado nivel  $p$ .

La riqueza final para los no accidentados si no se realiza fraude viene dada por  $W_f = W - l_i$ , la riqueza final de los que son descubiertos realizando fraude será de  $W_f = W - B - l_i - D_i$  mientras que la riqueza final de aquellos que no son descubiertos vendrá dada por  $W_f = W - D_i$ . Por último la utilidad de los prisioneros está representada por una función Von Neumann-Morgenstern  $U(\cdot)$  doblemente diferenciable.

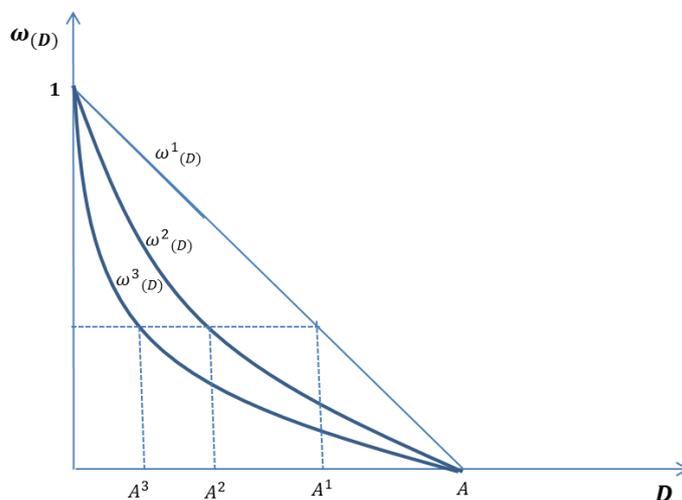
Se plantea un juego de tres etapas con dos variables estratégicas para los prisioneros. Primero la naturaleza decide si el prisionero sufrirá un ataque alto o bajo ( $\sigma$ ) y si sufrió un accidente ( $\delta$ ). En segundo lugar todos los prisioneros reclaman pago si sufren un accidente. Los que no sufren accidente cometen fraude con probabilidad ( $\alpha$ ) y deciden autolesionarse un nivel ( $D$ ). Luego de que los accidentes fueron reportados en la etapa 2, los carceleros auditan con probabilidad ( $p$ ) y descubren fraude con probabilidad  $p\omega_{(D)}$ . Se asume que el carcelero puede comprometerse a una estrategia de auditoria

por lo que elige la probabilidad de auditoría a partir de la función de reacción de los prisioneros no accidentados.

La función de engaño resume la capacidad de los prisioneros de disminuir la probabilidad de ser descubiertos al autolesionarse, por ejemplo aquellos con mejor capacidad para mentir podrán disminuir la probabilidad con mayor facilidad, por otro lado la función de engaño también refleja la capacidad de los carceleros de descubrir un fraude al observar un prisionero, carceleros ineptos podrán ser engañados con poco daño. Intuitivamente una tecnología de engaño decreciente y convexa implica que con relativamente poco daño puede caer rápidamente la probabilidad efectiva de ser detectado. Aquellas con mayor convexidad garantizan un mismo nivel de engaño con menor autolesión.

Es posible pensar en un caso extremo donde una cantidad ínfima de daño garantice no ser detectado  $\omega_{(D \approx 0)} = 0$  lo cual ocurriría si el prisionero es extremadamente habilidoso engañando o el carcelero es un completo inepto detectando fraude. El gráfico I muestra tres funciones. Se puede considerar a la función  $\omega^3_{(D)}$  más eficiente que las otras ya que para producir la misma cantidad de engaño, necesita menor cantidad de daño.

**Gráfico I – Funciones de engaño**



Aquellos que sufrieron un accidente no deben tomar ninguna decisión una vez que conocen su condición, al ser trasladados y no sufrir ataque su utilidad será  $U(W - A)$ . La utilidad esperada de los prisioneros que no sufrieron accidente pero que serán atacados vendrá dada por (9).

$$EU_i = \alpha_i [p\omega_{(D,A)}U_{(W-B-D-l_i)} + (1 - p\omega_{(D,A)})U_{(W-D)}] + (1 - \alpha_i)U_{(W-l_i)} \quad (9)$$

La autolesión genera por un lado una pérdida cierta de utilidad pero también una disminución de la probabilidad efectiva de ser descubierto. La utilidad esperada puede aumentar si para niveles relativamente bajos de daño se garantiza una caída más que proporcional de la probabilidad efectiva de ser descubierto y por ende castigado.

Tomando la primera derivada con respecto a  $\alpha$  se encuentra la utilidad esperada del prisionero si decide hacer fraude para un nivel de auditoría dado.

$$\frac{\partial EU}{\partial \alpha_i} = p\omega_{(D)}U_{(W-B-D-l_i)} + (1 - p\omega_{(D)})U_{(W-D)} - U_{(W-l_i)} \quad (10)$$

Adicionalmente el asegurado podrá autolesionarse para aumentar su utilidad esperada, dicha condición implica encontrar un nivel de daño óptimo al derivar la condición (10) con respecto a  $D$  lo que permite encontrar la condición de primer orden (11)

$$\frac{\partial EU_{(p,D)}}{\partial \alpha_i \partial D} = p \left\{ \frac{\partial \omega_{(D)}}{\partial D} [U_{(W-B-D-l_i)} - U_{(W-D)}] + \omega_{(D)} [U'_{(W-D)} - U'_{(W-B-D-l_i)}] \right\} - U'_{(W-D)} = 0 \quad (11)$$

A partir de la ecuación (11) y el cumplimiento de la condición de segundo orden presentada en el anexo se encuentra un nivel de autolesión que maximiza la utilidad esperada. El daño óptimo surge entonces de una interacción entre el costo cierto de autolesionarse y la ganancia esperada por la disminución de la probabilidad efectiva de ser descubierto. Sin embargo la estrategia óptima del prisionero será cometer fraude

mientras la condición (10) sea positiva para un dado nivel de auditoría y de daño óptimo, a continuación se formalizará dicho razonamiento.

**Resultado 1.** Dado un nivel de auditoría  $p > 0$  existirá al menos una tecnología de engaño  $\omega_{(A,D)}$  suficientemente convexa que garantice un nivel de daño  $D_{(p)}^* \in (0, A]$  que maximiza la utilidad esperada de hacer fraude.

Demostración: Anexo

El resultado 1 sugiere que si la tecnología de simulación es lo suficientemente eficiente para disminuir rápidamente la probabilidad de ser detectado entonces existirán niveles de auditoría para los cuales la estrategia óptima de los prisioneros será autolesionarse. Se observa en la ecuación (10) que si no hay auditoría no tiene sentido lesionarse. Sin embargo a medida que la probabilidad de auditoría aumenta, autolesionarse puede aumentar la utilidad esperada.

El proceso de maximización es el siguiente, dado un nivel de auditoría el prisionero determinará el nivel de autolesión que maximiza su utilidad esperada<sup>10</sup> a partir de (11).

Una vez que se conoce el daño óptimo  $D_{(p)}^*$  para un dado nivel de auditoría, el prisionero debe preguntarse si vale la pena realizar fraude, en términos del modelo esto implica conocer el comportamiento de la condición de maximización (10), si se cumple que  $\frac{\partial EU_{(p,D^*)}}{\partial \alpha_i} > 0$  entonces el prisionero decidirá hacer fraude, si  $\frac{\partial EU_{(p,D^*)}}{\partial \alpha_i} \leq 0$  entonces no cometerá fraude.

La estrategia óptima de los prisioneros vendrá dada por (12), estos tendrán incentivos a realizar fraude en la medida que la probabilidad de ser detectados esté por debajo de una probabilidad limite que se denominará  $p_i^*$ . Dado que  $l_b < l_a$  entonces  $p_b^* < p_a^*$  por lo

---

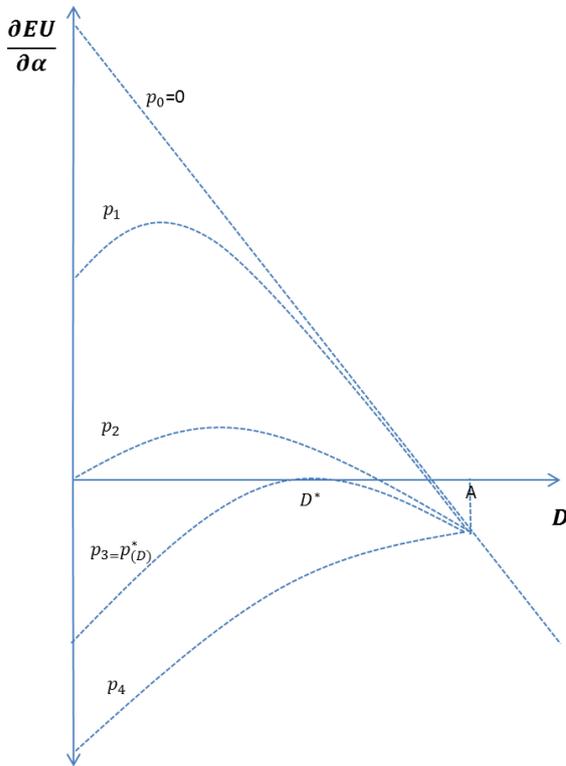
<sup>10</sup> El nivel de daño óptimo nunca será mayor que el daño que se intenta simular ya que se supuso  $\omega_{(D \geq A)} = 0$ .

que una auditoría que desincentive a los de ataque alto también desincentivará a los de ataque bajo.

$$\alpha_{(D^*)} \begin{cases} = 0 \text{ si } \frac{\partial EU_{(p,D^*)}}{\partial \alpha} < 0 \text{ con } p > p_i^* \\ \in [0,1] \text{ si } \frac{\partial EU_{(p,D^*)}}{\partial \alpha} = 0 \text{ con } p = p_i^* \\ = 1 \text{ si } \frac{\partial EU_{(p,D^*)}}{\partial \alpha} > 0 \text{ con } p < p_i^* \end{cases} \quad (12)$$

El gráfico II resume el comportamiento de  $\frac{\partial EU_{(p,D)}}{\partial \alpha}$  con respecto a la autolesión para distintos niveles de auditoría  $p_0 < p_1 < p_2 < p_3 < p_4$  si  $A < l_i$ . Como  $p_0 = 0$  entonces el daño será nulo ya que no existirán beneficios por autolesionarse, si el nivel de auditoría es  $p_1$  entonces existirá un nivel de daño que maximiza (11) y que garantiza que (10) sea positivo por lo que a partir de (12) se concluye que el prisionero realizará fraude. Para  $p_2$  se observa que ante la ausencia de daño la utilidad esperada es nula pero existen niveles de daño que la hacen positiva.

**Gráfico II – Incentivos a cometer fraude autolesionándose si  $A > l_i$**



El nivel de auditoría  $p_3 = p_i^*$  garantiza que con daño óptimo la ecuación (10) es nula por lo que a partir de (12) se concluye que  $\alpha_i \in [0,1]$ . Si el nivel de auditoría es  $p_4$  existirá un máximo en  $D_{(p_4)}^* = A$ . Se observa además que  $\frac{\partial EU}{\partial \alpha_i} < 0$  para  $D^* = A$  a ya que se supone que  $A > l_i$ . Para valores  $D^* > A$  la ecuación (10) será decreciente para cualquier nivel de auditoría debido a que  $\omega_{(D \geq A)} = 0$  por lo que no se obtendrá ningún beneficio adicional por lesionarse para valores  $D^* > A$ .

**Resultado 2.** Si  $A = l_i$ ,  $\frac{\partial EU_{(p,D^*)}}{\partial \alpha} = 0$  si y solo si existe un nivel  $p \geq p_i^* \in [0,1]$  que garantice  $D^* = l_i$  de modo que  $\omega_{(D^*=l_i)} = 0$ .

**Resultado 3.** Si  $A > l_i$ ,  $\frac{\partial EU_{(p,D^*)}}{\partial \alpha} = 0$  si y solo si existe un nivel  $p = p_i^* \in [0,1]$  que garantice  $D^* < l_i$  de modo que  $\omega_{(D^*)} > 0$ .

Demostración: Anexo

Estos dos ayudarán en la derivación de resultados posteriores, nótese que si el accidente simulado representa una pérdida mayor que el ataque futuro ( $A > l_i$ ) entonces el prisionero nunca se lesionará completamente, mientras que si son exactamente iguales el nivel de auditoría que desincentiva el fraude garantizará que el daño auto infligido será exactamente igual a la lesión que se desea evitar.

**Resultado 4.** Si  $A < l_i$  entonces aún con daño total es decir  $D_{(p)}^* = A$ , se cumplirá que

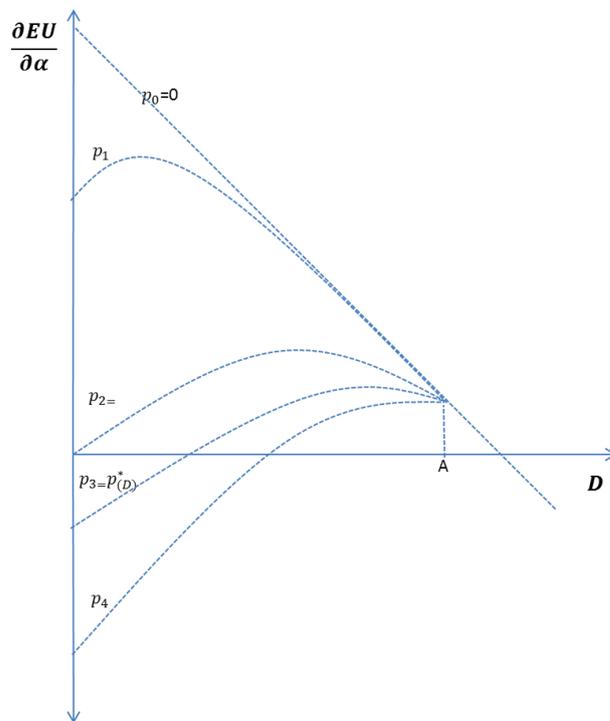
$\frac{\partial EU_{(p,D^*)}}{\partial \alpha} > 0$  para cualquier nivel de auditoría.

Demostración: Anexo

El resultado 4 representa la primera implicancia fuerte del modelo, aun si auditaran todos los casos el carcelero nunca podrá demostrar efectivamente que la lesión es un fraude si la lesión es total. Es importante notar que este resultado no depende de la convexidad de la tecnología de engaño si no del supuesto  $\omega_{(D \geq A)} = 0$ , esto quiere decir

que si  $D^* = A$  el carcelero no cuenta con las herramientas suficientes para catalogar el caso como fraude. Por ejemplo si un prisionero sabe que va a ser asesinado ( $W - l_i = 0$ ), cualquier simulación con autolesión que le garantice evitar la muerte será beneficiosa. El gráfico III muestra que ocurre si se supone  $A < l_i$ , el punto de convergencia se encuentra en la zona  $\frac{\partial EU}{\partial \alpha_i} > 0$ . Puede demostrarse a partir de (14) que si  $A = l_i$  entonces el punto de convergencia se encontrará en  $\frac{\partial EU}{\partial \alpha} = 0$ .

**Gráfico III - Incentivos a cometer fraude autolesionándose si  $A < l_i$ .**



**Resultado 5.** Aun si  $A > l_i$  no necesariamente existirá un nivel de auditoría  $p_i^* \in [0,1]$

que garantice  $\frac{\partial EU(p, D^*)}{\partial \alpha} = 0$

Demostración: Anexo

El resultado 5 representa la segunda implicancia fuerte del modelo. Si se define  $p_i^*$  como la probabilidad límite que garantiza que el prisionero no cometa fraude, esta no necesariamente se encontrará en el intervalo  $[0,1]$ . Matemáticamente esto implica que

podría ocurrir que la probabilidad límite sea mayor que la unidad lo cual resultaría un absurdo. La ecuación (13) representa la probabilidad efectiva que desalienta el fraude la cual si cumple la condición  $p_i^* \omega_{(D^*)} \in (0,1)$  con  $\omega_{(D^*)} > 0$ .

$$p_i^* \omega_{(D^*)} = \frac{U_{(W-D^*)} - U_{(W-l_i)}}{U_{(W-D^*)} - U_{(W-B-l_i-D^*)}} \in (0,1) \quad (13)$$

Sin embargo no hay motivos que garanticen que  $p_i^* \in [0,1]$  ya que  $\omega_{(D)} \in (0,1)$ . En el gráfico II puede observarse un caso particular donde  $p_3 = p_i^*$  garantiza  $\frac{\partial EU_{(p,D^*)}}{\partial \alpha} = 0$  sin embargo nada asegura a priori que la probabilidad de auditoría exista en el intervalo  $p_i^* \in [0,1]$ .

**Resultado 6.** Si  $A = l_i$ , entonces existe un nivel de auditoría  $\bar{p}$  que maximiza la probabilidad efectiva de descubrir un fraude.

Demostración: Anexo

El resultado 6 tiene una fuerte implicancia ya que valores altos de auditoría pueden implicar menor probabilidad efectiva de descubrir un simulador. Suponiendo que los prisioneros siempre defraudan entonces  $p = 0$  y  $p = p_{(D^*)}^*$  implican que la probabilidad efectiva será nula ya que  $p_i^* \omega_{(D^*)} = 0$ , para  $p < p_{(D^*)}^*$  se cumplirá que  $\frac{\partial EU_{(\bar{p},D^*)}}{\partial \alpha} > 0$  lo que implica  $\alpha = 1$ .

La probabilidad efectiva nunca será negativa por lo que un camino progresivo hacia  $p_{(D^*)}^*$  conllevará la existencia de un máximo  $\bar{p}$  y luego una disminución hasta el punto que sean indiferentes entre cometer fraude o no, es decir, donde  $\frac{\partial EU_{(p^*,D^*)}}{\partial \alpha} = 0$  y por ende  $\alpha = 0$ . Si  $A \leq l_i$  puede existir un nivel  $\bar{p} < p_{(D^*)}^*$  pero para simplificar el análisis se considerará solo el caso  $A = l_i$ .

Por su parte los carceleros buscaran minimizar sus costos, la probabilidad efectiva de descubrir un prisionero deshonesto afectará el costo esperado por el esfuerzo.

$$EC = \delta e + (1 - \delta)e[(1 - p\omega_a)\alpha_a\sigma + (1 - p\omega_b)\alpha_b(1 - \sigma)] \quad (14)$$

$$AC = \delta pc + (1 - \delta)pc[\alpha_a\sigma + \alpha_b(1 - \sigma)] \quad (15)$$

Donde (14) es el costo esperado por el esfuerzo de trasladar accidentados, el primer término corresponde al esfuerzo por ayudar prisioneros accidentados mientras que el segundo término indica el esfuerzo por ayudar a prisioneros simuladores no detectados en el proceso de auditoría. La ecuación (15) es el costo esperado de la auditoria, el primer término se refiere al costo de auditar prisioneros efectivamente accidentados mientras que el segundo representa el costo de auditar prisioneros no accidentados. El costo total será  $C = EC + AC$

Los carceleros elegirán el nivel de auditoría  $p$  que minimice sus costos totales, a los fines de encontrar un equilibrio tres niveles serán de interés. Para la obtención del equilibrio se definirá  $C_n$  en el caso particular de que no haya auditoría,  $C_b$  si solo realizan fraude los de ataque alto y por último  $C_a$  si se desincentiva completamente el fraude.

**Equilibrio 2.** Si existe compromiso a una política de auditoría y se cumple que  $A \leq l_i$  y existe  $p_i^* \in [0,1]$  el equilibrio de un juego de auditoría está caracterizado por

$$\begin{aligned}
 & p = 0, \alpha_a = \alpha_b = 1 \text{ y } D_a = D_b = 0 \text{ si } C_n \text{ es mínimo} \\
 & p = p_b^*, \alpha_a = 1 \alpha_b = 0 \text{ y } D_a = D_a^* D_b = 0 \text{ si } C_b \text{ es mínimo} \\
 & p = p_a^*, \alpha_a = \alpha_b = 0 \text{ y } D_a = D_b = 0 \text{ si } C_a \text{ es mínimo} \\
 & \min\{C_n, C_b, C_a\}
 \end{aligned}$$

Si  $A > l_i$  o si  $p_i^* \notin [0,1]$  entonces el equilibrio de un juego de auditoría está caracterizado por

$$p = 0, \alpha_a = \alpha_b = 1 \text{ y } D_a = D_b = 0$$

Demostración: Anexo

En primer lugar por resultado 4 la pérdida futura  $l_i$  necesariamente debe ser menor al accidente que simulan, si no es así entonces el prisionero siempre podrá lesionarse más y obtener un beneficio. En segundo lugar, por resultado 5 se deduce que la probabilidad de auditoría debe caer dentro del rango factible, si no es así aún una auditoría total no detendrá a los deshonestos. Por resultado 6 el comportamiento de los costos no es lineal con respecto a la auditoría por lo que es posible que existan equilibrios para valores de  $\bar{p}$  intermedios sin embargo en el anexo se demostrará que esto no es posible.

## Parte V. Comparación entre modelo *screening* y modelo de fraude.

Suponiendo funciones de utilidad lineales y  $\omega_{(D,A)} = 1 - \left(\frac{D}{A}\right)^\varphi$  resulta sencillo encontrar el nivel óptimo de daño para un dado nivel de auditoría a partir de condición de primer orden (11). El parámetro  $\varphi$  determina la convexidad de la función, si  $\varphi \in (0,1)$  entonces la función  $\omega_{(D)}$  será decreciente y convexa, cuanto menor sea el parámetro, mayor será la facilidad de los prisioneros para disminuir la probabilidad de auditoría.

$$D^* = \left(\frac{p(l_i+B)\varphi}{A^\varphi}\right)^{\frac{1}{1-\varphi}} \quad (16)$$

La ecuación (16) muestra que el daño óptimo será mayor cuanto mayor sea el castigo por ser descubierto, la probabilidad de ser auditado y la pérdida futura. Esto se debe a que un aumento de alguno de estos parámetros genera mayores pérdidas al ser detectado, por lo que es necesario aumentar el daño para disminuir la probabilidad de ser detectado.

El comportamiento  $\frac{\partial D^*}{\partial A} < 0$  merece atención, la función de engaño indica que cuanto más grave es el accidente simulado, mayor daño se necesitará para conseguir el mismo nivel de engaño, es decir si  $A_1 > A_2$  entonces  $\omega_{(D,A_1)} > \omega_{(D,A_2)}$ , sin embargo la pérdida cierta por autolesión en (9) impactará igual independientemente de la gravedad del accidente simulado por lo que simular lesiones más graves desincentiva la autolesión ya que disminuyen los beneficios.

No puede obtener un resultado general de  $\frac{\partial D^*}{\partial \varphi}$ , si el denominador de la ecuación (16) es mayor que el numerador entonces existirá un valor de  $\varphi$  que maximiza el daño, Intuitivamente, una función de engaño menos eficiente implica que el costo de autolesionarse será relativamente más alto ya que disminuye el ingreso generado por la disminución de la probabilidad efectiva de ser detectado, por ende el daño óptimo podría ser menor para compensar la disminución de los ingresos.

Si se supone  $A = l_i$  entonces utilizando el resultado 2 y (16) es posible hallar el valor  $p_a^*$  que desincentiva el fraude. Puede observarse en (17) que para valor bajos de  $\varphi$  es posible que la probabilidad no se encuentre en el rango factible, lo que confirma el resultado 5 para funciones de utilidad lineales. Además se observa que cuenta mayor el castigo menor será la probabilidad límite.

$$p_a^* = \frac{l_a}{(l_a+B)\varphi} \quad (17)$$

El equilibrio 2 muestra que existirá autolesión si los carceleros deciden auditar hasta el nivel que desincentive el fraude de los prisioneros que sufrirán un ataque bajo. A los fines de realizar la comparación con el modelo de *screening* se encontrará la utilidad esperada de los prisioneros que sufrirán un ataque alto suponiendo funciones de utilidad lineales y  $\alpha_b = 0$ .

$$EU = l_a - D^* - p_b^* \omega^* (l_a + B) \quad (18)$$

A partir de la ecuación (9) se obtiene la utilidad esperada (18) la cual representa la diferencia entre la pérdida futura que se evita  $l_a$  y dos términos, el primero de ellos representa la pérdida cierta por autolesión, el segundo de ellos es la pérdida por ser detectado lo que implica el castigo  $B$  y el ataque futuro que no podrá ser evitado.

Hasta ahora se supuso que lo que incentiva el fraude es la ausencia de beneficios para los carceleros ya que esto impedía la existencia de un equilibrio separador al no cumplirse las restricciones de participación (3) y (4). Sin embargo el juego de fraude podría ser preferido aún si los carceleros tuviesen una ganancia por evitar la pérdida futura de un prisionero con peligro alto, es decir si  $a_{(l_a)} = a^* > 0$ . Ahora el carcelero recibe beneficios adicionales si impide un ataque alto, pero también está obligado a atender a aquellos prisioneros accidentados sin recibir un beneficio por ello.

Al comparar el resultado (8) con el (18) se observa que en ambos casos la ganancia viene dada por la pérdida  $l_a$  que se evita, por lo que la condición de indiferencia (19) surge de igualar los costos de autolesionarse.

$$l_b = \left( \frac{p_b^* (l_a + B) \varphi}{A \varphi} \right)^{\frac{1}{1-\varphi}} + p_b^* \omega^* (l_a + B) \quad (19)$$

Por ende si el prisionero que sufrirá un ataque alto es relativamente bueno engañando, o la auditoría  $p_b^*$  o el castigo  $B$  son relativamente bajos entonces el fraude será preferido a un pedido de auxilio costoso y honesto. Una pérdida futura  $l_a$  más baja también incentiva el fraude ya que en caso de ser detectado la pérdida será menor independientemente de que los beneficios sean menores en ambos casos.

También puede ocurrir que exista una correlación entre los beneficios del jugador no informado y el castigo si descubre fraude. Para ejemplificar, un carcelero muy protector y considerado de los prisioneros que sufrirán un ataque alto está dispuesto a ayudarlos si prueban su condición, a su vez tampoco lo castigaría si son descubiertos simulando una lesión. Esta asimetría en el castigo entre jugadores de peligro alto y bajo plantea una muy interesante paradoja ya que cuanto más protector y benevolente es el carcelero, menor será el castigo por lo que mayores serán los incentivos del prisionero a cometer un fraude y menores los incentivos a ser honesto.

## **Parte VI - Aspectos complementarios**

Un juego típico de *Screening* presenta tres condiciones necesarias. En primer lugar el jugador informado realiza sus acciones en respuesta a un contrato de protección prefijado por el no informado. En segundo lugar existe una acción que el jugador no informado puede realizar y que beneficia al informado independientemente de su naturaleza. Por último la acción beneficiará al no informado si y solo si el otro jugador realmente se encuentra en una situación de peligro. Existe entonces la posibilidad de que ambos jugadores se beneficien si el verdadero estado de la naturaleza es revelado por la acción costosa del jugador informado.

En cambio el juego de fraude presentado no necesita de la tercera condición, de hecho requiere que el jugador no informado este obligado ya sea por contrato o por altos

costos de reputación (por ejemplo exposición en los medios de comunicación, Boin y Rattray 2004) a responder solamente ante la ocurrencia de un estado accidental de la naturaleza. Incluso la literatura criminológica (Hepburn 1985) menciona que puede ocurrir que el carcelero decida castigar al prisionero si este se autolesiona, por ende el fraude se presenta como una buena alternativa.

Un requisito previo tanto para los modelos de *screening* como para los de fraude es la imposibilidad de los jugadores informados de elegir o cambiar al jugador no informado (Hagen et al 2008), no existe un mercado de carceleros benevolentes ni de madres cariñosas por lo que en muchos casos una alternativa eficiente a la autolesión sería el soborno o la corrupción, por ejemplo el prisionero podría recurrir a un médico cómplice para obtener un certificado médico que garantice un traslado pero esta opción dependerá de características propias de cada jugador.

Alternativamente podría darse el caso de que ambos jugadores no estén informados de la condición del otro, por ejemplo si el prisionero no sabe con certeza si el carcelero tendrá ingresos adicionales por detener un ataque antes de que ocurra es posible que la estrategia del fraude domine a la de *screening* ya que en ambos casos obtendrá un beneficio.

La autolesión como herramienta de fraude no solo evita pérdidas futuras sino que también puede reportar ingresos, por ejemplo un trabajador puede autolesionarse para cobrar el seguro laboral o un político puede simular una suba de presión para hacer notar el rechazo que siente hacia la postura de la oposición y lo comprometido que está con su discurso.

Este trabajo demostró que la autolesión como herramienta de fraude puede servir como alternativa a un juego de pedido de auxilio, sin embargo no será una alternativa a todo

tipo de juego de señales o *screening*. Por ejemplo si la autolesión es señal de compromiso en un rito de iniciación entonces simular un estado accidental de la naturaleza no tendrá mucho sentido, tampoco lo tendrá cuando se trata de demostrar fuerza para detener atacantes, aunque puede darse el caso de que exista reglas implícitas del tipo “no pelear si la otra persona está accidentada” que impida un ataque y fomente la simulación.

## **Parte VII. Conclusión**

Es claro que la autolesión puede ser utilizada para distintos fines, la literatura sociológica y biológica considera modelos de autolesión bajo la óptica señales mientras que la literatura psicológica se concentra principalmente en motivos intrapersonales. Este trabajo en cambio propone modelar el comportamiento de individuos que deciden realizar fraude. Este enfoque permite demostrar que la simulación es una alternativa válida a juegos de información asimétrica de tipo *Screening*. El fraude puede ser una alternativa viable aun cuando el jugador no informado obtiene beneficios si la verdad se transmite. Por lo que futuros estudios empíricos sobre autolesión podrían ser tratados bajo metodologías propias de la teoría del fraude.

El modelo de fraude presentado considera que la autolesión es un mecanismo para disminuir la probabilidad de ser descubierto por lo que se aparta de los modelos tradicionales de verificación y falsificación costosa. El suponer una función de engaño permite concluir que aun si todos los casos son auditados aún pueden existir incentivos para defraudar. Además, si la lesión simulada representa una pérdida menor que el ataque futuro entonces no existirá auditoría suficiente que evite la autolesión.

## Bibliografía

- Becker, G. S., & Elias, J. J. (2007). Introducing incentives in the market for live and cadaveric organ donations. *The Journal of Economic Perspectives*, 3-24.
- Bentley, K. H., Nock, M. K., & Barlow, D. H. (2014). The Four-Function Model of Nonsuicidal Self-Injury Key Directions for Future Research. *Clinical Psychological Science*, 2(5), 638-656.
- Boin, A., & Rattray, W. A. (2004). Understanding Prison Riots Towards a Threshold Theory. *Punishment & Society*,
- Bond, E. W., & Crocker, K. J. (1997). Hardball and the soft touch: the economics of optimal insurance contracts with costly state verification and endogenous monitoring costs. *Journal of Public Economics*, 63(2), 239-264.
- Bourgeon, J. M., & Picard, P. (2000). Reinstatement or insurance payment in corporate fire insurance. *Journal of Risk and Insurance*, 507-526.
- Crocker, K. J., & Morgan, J. (1998). Is honesty the best policy? Curtailing insurance fraud through optimal incentive contracts. *Journal of Political Economy*, 106(2), 355-375.
- Dahle, K. P., Lohner, J. C., & Konrad, N. (2005). Suicide prevention in penal institutions: Validation and optimization of a screening tool for early identification of high-risk inmates in pretrial detention. *International Journal of Forensic Mental Health*, 4(1), 53-62.
- Dawkins, R. (2006). *The selfish gene*. Oxford university press.
- Gambetta, D. (2009). *Codes of the underworld: How criminals communicate*. Princeton University Press.
- Gambetta, D. (2009b). Signaling. *The Oxford Handbook of Analytical Sociology*, 168-194.
- Griller, J. (2013). *The Paradox of Self-harm in Prison: Psychopathy Or an Evolved Coping Strategy?*. Diplomica Verlag.
- Hagen, E. H., Watson, P. J., & Hammerstein, P. (2008). Gestures of despair and hope: A view on deliberate self-harm from economics and evolutionary biology. *Biological Theory*, 3(2), 123.
- Heide, S., & Kleiber, M. (2006). Selbstbeschädigung—eine rechtsmedizinische Betrachtung. *Dtsch Arztebl*, 103(40), 2627-33.
- Hepburn, J. R. (1985). The exercise of power in coercive organizations: a study of prison guards\*. *Criminology*, 23(1), 145-164.
- Kaminski, M. (2004). *Games prisoners play: The tragicomic worlds of Polish prison*. Princeton NJ: Princeton University Press.
- Klonsky, E. D., Muehlenkamp, J. J., Lewis, S. P., & Walsh, B. (2011). *Nonsuicidal self-injury*. Hogrefe Pub..
- Nock, M. K. (2008). Actions speak louder than words: An elaborated theoretical model of the social functions of self-injury and other harmful behaviors. *Applied and Preventive Psychology*, 12(4), 159-168.
- Nock, M. K. (2009). *Understanding nonsuicidal self-injury: Origins, assessment, and treatment*. American Psychological Association.
- Nock, M. K., & Prinstein, M. J. (2004). A functional approach to the assessment of self-mutilative behavior. *Journal of consulting and clinical psychology*, 72(5), 885.
- Picard, P. (1996). Auditing claims in the insurance market with fraud: The credibility issue. *Journal of Public Economics*, 63(1), 27-56.
- Picard, P. (2013). Economic analysis of insurance fraud. In *Handbook of Insurance* (pp. 349-395). Springer New York.
- Rasmusen, E. (2007). *An introduction to game theory*. Blackwell Publishing
- Roth, A. E. (2006). *Repugnance as a Constraint on Markets* (No. w12702). National Bureau of Economic Research

## ANEXO

### Demostración resultado 1.

Se renombrará  $U_{(\cdot)}^M = U_{(W-B-D-l_i)}$  y  $U_{(\cdot)}^B = U_{(W-D)}$ . Se sabe que  $0 < U_{(\cdot)}^M < U_{(\cdot)}^B$ ;  $U_{(\cdot)}^M > U_{(\cdot)}^B > 0$  y  $U_{(\cdot)}^{''M} < U_{(\cdot)}^{''B} < 0$ , además  $\frac{\partial \omega}{\partial D} < 0$  y  $\frac{\partial^2 \omega}{\partial^2 D} > 0$ , la condición de segundo orden se obtiene a partir de derivar (15) con respecto al daño

$$\frac{\partial EU}{\partial \alpha \partial^2 D} = p \left\{ \frac{\partial^2 \omega}{\partial^2 D} [U_{(\cdot)}^M - U_{(\cdot)}^B] + 2 \frac{\partial \omega}{\partial D} [U_{(\cdot)}^{'B} - U_{(\cdot)}^{'M}] + \omega [U_{(\cdot)}^{''M} - U_{(\cdot)}^{''B}] \right\} + U_{(\cdot)}^{''B} \quad (20)$$

El primer, tercer y cuarto término son negativos mientras que el segundo término es positivo por lo que nada a priori garantiza que se cumpla la condición de segundo orden. Sin embargo puede pensarse en una función de engaño muy eficiente que con niveles de autolesión cercanos a cero garanticen que  $\omega_{(A,D)} \cong 0$  por lo que una cantidad pequeña de daño garantizará un aumento en (10).

El análisis se simplifica ara funciones de utilidad lineales ya que  $\frac{\partial EU}{\partial \alpha \partial D} = -p \frac{\partial \omega}{\partial D} [l_i + B] + 1 = 0$  y la condición de segundo orden  $\frac{\partial EU}{\partial \alpha \partial^2 D} = -\frac{\partial^2 \omega}{\partial^2 D} [l_i + B]$  será negativa debido al supuesto de convexidad de la función de engaño.

### Demostración resultado 2, 3 y 4.

Resultado 2. Dado  $p_i^*$  se encuentra en (11) el nivel de daño  $D^* = A$ , entonces (10) será igual a cero ya que se supone  $A = l_i$  y  $\omega_{(A,D)} = 0$ . El resultado  $\frac{\partial EU_{(p,D=A)}}{\partial \alpha_i} = 0$  solo puede ocurrir para  $p_i^*$  ya que si  $p < p_i^*$  entonces el daño será menor a la lesión simulada y la utilidad esperada en (12) será positiva, notar que cualquier valor  $p > p_i^*$  la utilidad esperada de hacer fraude seguirá siendo cero ya que  $\omega_{(A,D)} = 0$  garantiza que (10) no dependa de la probabilidad de auditoría.

Resultado 3. Si  $p_i^*$  garantiza  $D^* = A$  entonces (10) será menor que cero ya que se supone  $A > l_i$  por lo que el valor de auditoría que garantiza que (10) sea cero tiene que tener un daño asociado  $D^* < A$  y por ende  $\omega_{(A,D)} > 0$ . El resultado  $\frac{\partial EU_{(p,D=A)}}{\partial \alpha_i} = 0$  solo puede ocurrir para  $p_i^*$  ya que si  $p < p_i^*$  entonces la utilidad esperada en (12) será positiva, si  $p > p_i^*$  la utilidad esperada de hacer fraude será negativa por los mismos motivos.

Resultado 4. El resultado puede demostrarse a partir de la ecuación (14) si se asume  $D = A$  lo que permite llegar a la ecuación  $\frac{\partial EU_{(p,D=A)}}{\partial \alpha_i} = U_{(W-A)} - U_{(W-l_i)} > 0$  si  $A < l_i$  para todo  $p$ .

#### **Demostración resultado 5.**

El razonamiento para hallar (17) es el siguiente, para un nivel de auditoría  $p_i^*$  se calcula a partir de (11) el nivel de autolesión  $D^*$ , igualando (10) a cero se despeja la probabilidad efectiva  $p_{(i)}^* \omega_{(D)} \in (0,1)$ . A partir de la ecuación (17) es evidente que si  $B > 0$  entonces la probabilidad efectiva  $p_i^* \omega_{(D)}$  será menor que la unidad, sin embargo nada garantiza que  $p_i^*$  se encontrará en el rango factible ya que  $\omega_{(D)} \leq 1$ .

#### **Demostración resultado 6.**

Si no existe auditoría entonces no existirán incentivos a autolesionarse, es decir  $p \omega_{(D^*=0)} = 0$  si  $p = 0$ , por otro lado si se supone por un momento que los individuos siempre defraudan ( $\alpha_i = 1$ ) entonces existirá un nivel  $\hat{p}$  (no necesariamente menor que uno) que garantizará daño total  $D^* = A$  por lo que  $\hat{p} \omega_{(D^*=A)} = 0$ . En el caso particular  $A = l_i$  se cumplirá que  $\hat{p} = p_i^*$ . Entre los dos extremos mencionados debe cumplirse que  $\omega_{(D^*)} \geq 0$  por lo que existirá un nivel de auditoría  $\bar{p} \in (0, \hat{p})$  que maximizará la

probabilidad efectiva  $p\omega_{(D^*)}$  si se cumple que  $\bar{p} < p_i^* \leq 1$  ya que si  $\bar{p} \geq p_i^*$  entonces  $\alpha_i = 0$ . Si  $A = l_i$  entonces el hecho de que  $\hat{p} = p_i^*$  garantizará  $\bar{p} < p_i^*$ .

### **Demostración equilibrio 1**

Debido al resultado 6 no puede asegurarse que los costos sean siempre decrecientes o crecientes con respecto al nivel de auditoría. Primero se encontrará las condiciones de equilibrio para  $p = 0$ ,  $p_b^*$  y  $p_a^*$  y luego se demostrará que no existe equilibrio en niveles intermedios. El costo total si se supone  $\alpha_i = 0$  ocurrirá cuando  $p \geq p_a^*$  y será mínimo cuando  $p = p_a^*$ . Dicho costo se denominará  $C_a$  y vendrá dado por (21)

$$C_a = \delta e + \delta c p_a^* \quad (21)$$

Ahora se considera el caso  $\alpha_a = 1$  y  $\alpha_b = 0$  que ocurrirá en el intervalo  $p_b^* \leq p \leq p_a^*$ .

Si  $p = p_a^*$  entonces el costo total será mayor que (21) pero nada puede decirse a priori para valores intermedios. El costo asociado si  $p = p_b^*$  vendrá dado por  $C_b$  en la ecuación (22). Luego se demostrará que para otros valores de auditoría no se minimizan los costos.

$$C_b = \delta e + \delta p_b^* c + (1 - \delta) \sigma [(1 - p_b^* \omega_{a(p_b^*)}) e + p_b^* c] \quad (22)$$

En el caso  $\alpha_i = 1$  ocurrirá si  $p \in [0, p_b^*]$ . Si  $p = p_b^*$  entonces el costo será mayor que (22). Si no hay auditoría el costo asociado será igual a (23).

$$C_n = \delta e + (1 - \delta) e \quad (23)$$

El costo de auditoría  $c$  que hace indiferente  $C_a$  y  $C_n$  vendrá dado por  $c_{an} = \frac{(1-\delta)e}{\delta p_a^*}$ , el

que hace indiferente  $C_a$  y  $C_b$  será  $c_{ab} = \frac{(1-\delta)e\sigma(1-p_b^*\omega_a)}{\delta p_a^* - p_b^*[\delta + (1-\delta)\sigma]}$  mientras que la indiferencia

entre  $C_b$  y  $C_n$  vendrá dada por  $c_{bn} = \frac{(1-\delta)e[1-\sigma(1-p_b^*\omega_a)]}{p_b^*[\delta + (1-\delta)\sigma]}$ . El costo total  $C_n$  será mínimo

si  $c > c_{an}$  y  $c > c_{bn}$ . Por su parte el costo total  $C_a$  será mínimo si  $c < c_{an}$  y  $c < c_{ab}$ . Mientras que  $C_b$  será mínimo si  $c \in (c_{ab}, c_{bn})$ . Para que esto último ocurra debe cumplirse que el denominador de  $c_{ab}$  sea positivo y que  $c_{ab} < c_{bn}$ . Si alguna de estas dos condiciones no se cumplen entonces  $C_b$  nunca será mínimo

Ahora se demostrará que en el caso  $\alpha_a = 1$  y  $\alpha_b = 0$  no se minimizarán los costos para valor de auditoría intermedios. De forma similar a (22) pueden encontrarse el costo total  $C_p$  para valores de auditoría  $p \in (p_b^*, p_a^*)$ .

$$C_p = \delta e + \delta p c + (1 - \delta) \sigma [(1 - p \omega_{a(p)}) e + p c] \quad (23)$$

El costo de auditoría que hace indiferente a  $C_p$  y  $C_a$  vendrá dado por  $c_{ap} =$

$$\frac{(1-\delta)e\sigma(1-p\omega_a)}{\delta p_a^* - p[\delta + (1-\delta)\sigma]}$$

mientras que la indiferencia entre  $C_p$  y  $C_a$  viene dada por  $c_{pn} =$

$$\frac{(1-\delta)e\sigma(p\omega_{(p)} - p_b^* \omega_{(p_b^*)})}{(p - p_b^*)[\delta + (1-\delta)\sigma]}$$

El costo total  $C_p$  será mínimo si el costo de auditoría  $c \in (c_{ap}, c_{pn})$  lo que implica que  $c_{ap} < c_{pn}$  pero puede demostrarse algebraicamente que esto nunca ocurrirá por lo que no existirá un valor de auditoría  $c$  que garantice que  $C_p$  sea el costo total mínimo. El mismo razonamiento puede utilizarse para  $\alpha_i = 1$  con niveles de auditoría asociados  $p \in (0, p_b^*)$ . Por ende los costos mínimos posibles serán  $\{C_a, C_b, C_n\}$ .