Markets for leaked information*

Steffen Huck       Georg Weizsäcker

August 13, 2015

Abstract

We study markets for sensitive personal information. An agent wants to communicate with another party but the revealed information can be intercepted and sold to a third party whose reaction harms the agent. The market for personal information induces an adverse sorting effect, such that the the sensitive information is allocated to those types of third parties who harm the agent most. In equilibrium, this limits information transmission by the agent, but never fully deters it. We also consider agents who naively provide information to the market. Their presence renders traded information more valuable and, thus, harms sophisticated agents by increasing the third party’s demand for information. Regulatory interventions that only modestly increase the cost of selling information in the market may hurt the naive agent without helping the sophisticated agents.

*We thank Micael Castanheira, Miguel Costa-Gomes, Ludwig Ensthaler, Drew Fudenberg, Philippe Jehiel, Sebastian Kodritsch, Dorothea Kübler, Stefan Lessmann, Roland Strausz, Daniel Warren and Jidong Zhou as well as audiences at Lund, NYU, Regensburg, UCL and WZB for helpful discussions. Financial support by the ERC (Starting Grant 263412) is gratefully acknowledged. Huck: WZB Berlin and University College London, Weizsäcker: Humboldt University Berlin and DIW Berlin. Contact: huck@wzb.eu, weizsaecker@hu-berlin.de.
1 Introduction

Some information travels far and often in ways that its originators fail to foresee. From their perspective, the information “leaks”; alternatively, an outside observer might call the senders “naive” for providing the information in the first place without taking care to communicate safely. New developments in information technology breed such naivety. Few people keep pace with the state of the art of data collection, data storage and data processing. Even fewer can anticipate what a future agent controlling some future technology will do with information that is recorded today. Yet many households appear not to worry and reveal sensitive information, especially in online social media and online commerce, without taking precautionary measures.\(^1\)

This paper investigates the market incentives arising from the possibility of leaked (or intercepted) personal information. We start with player A who has some information to share but faces a trade-off. On the one hand, she wants to share her information with player B because of some joint activity that benefits from truthful revelation. On the other hand, A fears the information ending up in the wrong hands—a player C who may take actions against A’s interest if he gets hold of her information. Such a trade-off in information revelation features in many economic applications. For example, a firm may want to share details of their technology with a supplier yet fears that competitors may get access to the information. Consumers may want to reveal their preferences at one shop yet prefer that the information is not shared with other shops to rule out price discrimination. Or, in a broader social realm, friends might want to exchange information about political, religious or sexual preferences yet may have reason to fear that this information is obtained by their employers, relatives or certain state agencies.

An important question is via what mechanism A’s information can end up in C’s hands. Our model studies the case where the information is traded in a market. There are two separate ways in which A’s data may be available for trade that our model captures. First, player B might engage in double dealing using the information fruitfully for his mutual business with A, but also selling it off for additional private gain to C. C’s preference for obtaining the information are unknown to A and B, with a distribution that generates a standard market demand curve. Second, and analogously, A’s information

\(^1\)See the evidence reported in Acquisti and Grossklags, 2005, Norberg et al., 2007, Tsai et al., 2011, and Beresford et al., 2012, as well as the survey article by Acquisti et al., 2014.
may intercepted by another player (not B himself) who sells it off to C.\footnote{For simplicity, we model this situation also as if it was B himself who sold the information on the market. As there are no complementarities or other spillovers between the two actions—using the information for generating joint surplus with A and selling it off—B’s utility will be separable in these two actions and can, thus, also capture the behavior of two different agents.}

Depending on the anticipated market allocation that arises from B’s and C’s interaction, Player A must be careful when communicating with B. A key variable for A to anticipate is the cost that B incurs for intercepting and selling the information. He will offer A’s data to C if its expected market revenue exceeds the investment cost required for data processing, including storage, formatting, refinement etc., as well as market transaction cost. In short, the cost reflects the technology that is available at the time of B’s market entry decision. We argue that this technology is difficult to assess. The typical consumer and user of online media may vastly underestimate the ease with which her data are captured, screened and evaluated. Even institutional users like firms or other (non-IT) professionals may have little information about the exact routing of their emails and the actual security of their data. A further reason for such biases is that the market entry decision may lie in the future, potentially years after the information revelation. Over time, advances in information technology drive down the cost of information processing and some of these advances are hidden from the affected users (and with few incentives for the informed parties to reveal them).\footnote{Similar statements apply if parts of the cost arise from regulatory compliance: the typical sender of information cannot know how costly data security regulation is or will be.} \footnote{If the seller is different from the recipient of A’s information, there are additional costs of obtaining the information be it through interception / hacking / stealing.}

We capture these issues by assuming that there are two types of player A: sophisticated types who know the true cost and naive types of player A who over-estimate the cost of selling information. In all other respects, our model assumes agents to be fully rational. In order to allow for such partial naivety, we formulate a game with non-common priors where both types of A agents are aware that there is another type with different beliefs about the cost of information provision.

The model makes several unambiguous equilibrium predictions. We hasten to say that it is “only a model” and that despite our attempts to make it general, it clearly rests on specific assumptions In our view, perhaps its most important simplifying assumption is that C’s and A’s incentives are oppo-
This assumption simplifies the incentives of player A, yields a unique symmetric equilibrium prediction and enables the full analysis for a large set of possible preference distributions.

This short paper makes, however, only a first step towards a fuller market analysis, by focusing on monopolistic market structure (Section 3.2) where player A reveals her information only to one, possibly two-faced, player B. Player B reacts to obtaining the information by choosing an action that helps A, but he also offers to C the sale of A’s information at a monopolistic price. If C obtains the information, he reacts with an action that suits himself and harms A. C’s reaction depends on the knowledge of his own type and of A’s type. Different types of C are more or less aggressive, but all types of C need to obtain information about A in order to identify their optimal reaction.

Our first result bears bad news for A: the monopolist sells the information to those types of C who harm A the most. This result follows from the assumption on opposite preference alignment of A and C. C’s willingness to pay for the information increases monotonically with the harm that he inflicts on A. The result is, of course, a rather generic property of markets: they allocate goods to those buyers with the highest willingness to pay—here, the most aggressive C types. But the harmful effect on A comes with a self-moderating equilibrium effect. Player A reacts to the danger of leakage by withholding information. The market traders B and C, in turn, are aware of A’s reluctance and react to it by exercising caution themselves: market activity will never be so intense as to completely shut player A up.

The market’s reaction to A’s withholding of information also highlights how differently the market traders benefit from trading information that originates from naive versus from sophisticated A’s. Sophisticated A correctly anticipates the market activity and withholds information whenever the equilibrium predicts the possibility of information leakage. The market traders therefore cannot earn much profit from her. Naive A, in contrast, is an easy victim and generates larger profits. For many parameter constellations, she provides strictly more information than sophisticated A, and the traders may even prefer to leave naive A fully in the dark about their trading.

5 The assumption is in contrast to much of the existing theoretical literature that we summarize below, but corresponds naturally to the description that A is worried about C and does not reveal the information voluntarily to him.

6 In the alternative interpretation of the model, where B’s information is intercepted, this corresponds to the case there is only one agent who can spy on A and sell the information.
activities. Indeed, we find that the parameter constellation where B’s profit from trading information is maximal is one where naive A feels uninhibited in her information provision while sophisticated A withholds information.

We also find an unambiguous effect of A’s possible naivety on her own payoff. Unsurprisingly, naive A receives smaller payoffs than sophisticated A. More surprisingly perhaps, we find a generally negative externality of naive A on sophisticated A. This externality arises because the prevalence of naivety increases the value of information, hence increases demand, and hence increases trading activity. Sophisticated A therefore needs to reduce information provision further, in response to naive A’s presence.

Ironically, a larger proportion of naive players may benefit a given naive player. As she realizes that players C can expect more information transmission with more players A who are like herself, she might become more cautious and hence better adjusted to the true costs of the monopolist. However, an increase in the degree of naivety (not its prevalence) makes both sophisticated and naive players A unambiguously worse off.

One possible consumer protection intervention is to educate consumers about the cost of obtaining and selling information. If such measures reduce the share of naive types, sophisticated As will benefit from such education.

Other possible regulatory interventions affect B’s cost of supplying information, e.g. through data security requirements or through bans of certain technologies. Our analysis shows how in equilibrium, such interventions may help A to provide information to B. We identify conditions under which all types of player A are better off if the cost of selling the information increases: if data protection is initially weak then a strong enough intervention helps to increase information provision for sure. However, we also show that ill-designed interventions—even if they do increase B’s costs—can cause harm without creating any benefit: they may encourage carelessness of naive types of A.

In Section 2 we discuss possible applications of our model in more detail and relate them to existing strands of the literature. Section 3 contains the formal analysis and Section 4 concludes.

2 Applications and literature

Business-to-business interaction. Player A is a manufacturer who requires some specific inputs such as software or machinery for which she has to share
detailed information about her technology with supplier B. This information is valuable to A’s competitors who might use it for their strategic advantage or to copy A’s technology. Similar problems arise with outsourcing such that, as Arrunada and Vazquez (2006) discuss in contexts of electronics and car manufacturing, A’s contract manufacturer can become her competitor. Ho (2009) has a related theoretical discussion on outsourcing showing that the presence of multiple upstream trading partners can alleviate the contracting problems that arise from the threat of information leakage.

**Social media and the workplace.** Player A is an employee who uses online social media (OSM) services with a friend or colleague, player B. In their conversations, A directly benefits from sharing details about her workplace situation with B. But her communication is recorded and may be offered on a market by a data aggregator who sieves A’s OSM usage data. The potential buyer, player C, is player A’s current or future employer who cares about A’s focus on her job or her views about the firm’s management. If A’s communication reveals some deviations from the employer’s expectations he can as the current employer redesign A’s job, redepoly or sack her. Future potential employers may think carefully about what type of job to offer to her (if any). Related evidence is in Acquisti and Fong’s (2014) large-scale field experiment showing some US employers indeed rely on OSM usage and that job market prospects are significantly affected through online revelation of personal traits.

**Personalized advertising and pricing.** Player A is a consumer whose choices are recorded through store cards or cookies. Player B is a seller who caters to A’s demands (generating surplus for both of them) but also collects the information about A’s type which he can sell to other companies.

There is a fast-growing theoretical and epirical literatures that investigate personalized advertisement and personalized pricing. For the theoretical literature see e.g. Taylor (2004), Armstrong and Zhou (2010), Jentzsch (2014), Bergemann and Bonatti (2015) and the survey on personalized pricing by the UK’s Office of Fair Trading (2013). For empirical evidence on firm behavior see e.g. the field experiment by Vissers et al (2014). In these literatures, it is a widespread view that consumers may not be aware of the possibility of subsequent data use and the allocation of information happens via market mechanisms. Evidence on consumer behavior that is (differentially) consistent with several kinds of consumer naivete is discussed in the experiments and surveys in Acquisti and Grossklags (2004), Norberg et al, (2007), Tsai

We point out to the reader that the cited theoretical papers do not describe special cases of our model. Our assumptions differ from those made in the theoretical literature on consumer privacy in at least two basic features. First, we focus on agents who only provide information and hence have no bargaining power and cannot achieve good deals in return for their information provision. Their only weapon is to withhold information—with the drawback of reducing the benefit from interacting with player B. Second, as discussed above, we focus on the case where the buyer of the information, player C, has incentives that are known to be oppositely aligned to those of player A. That is, his reaction can only harm A and the uncertainty about C pertains merely to the strength of his preference or ability to harm A.

Finally, we need to give credit to two theoretical literatures, even if only briefly. First, our model builds directly on the literature on strategic information transmission. The assumptions are closely related to those of Crawford and Sobel (1982) although the results are different. Second, the literature on contracting with differentially naive consumers (e.g. Gabaix and Laibson, 2006, Eliaz and Spiegler, 2006) has related assumptions on consumer naivety. Differently to most of this literature, we find that sophisticated agents do not benefit from others being naive but instead they unambiguously suffer from having naive counterparts.

3 The model

3.1 Main elements

We start the exposition with a summary of the model, leaving the details to the subsequent subsections.

Player A has private information $\theta_A \in \mathbb{R}$ that she wants to reveal to player B but not to player C. A sends a message $m_A$ to B, which is constructed such that when B reads it, he either learns the true $\theta_A$ or receives an empty message. Player A’s only choice variable in our model is the precision of her message to B: she picks the probability $x_A$ with which B learns $\theta_A$ from her message. Increasing the probability $x_A$ is costly and we assume that their cost $c_A \in \mathbb{R}$ is increasing, twice continuously differentiable and convex.\(^7\)

\(^7\)The cost reflects the effort that A has to spend to describe her information or to
Player B makes two decisions. On the one hand, he picks an action \( d_B \in \mathbb{R} \) for his joint project with player A. The action aims to match his beliefs about \( \theta_A \) and the more precisely it is matched to the truth, the more surplus it generates. On the other hand, a two-faced B can make an offer to player C, to sell him the message \( m_A \). This offer causes cost \( c_B \) to player B. Alternatively, we can think of there being another player, some B2, who intercepts A’s message and offers selling it to C. For notational simplicity we will stick to the unity of one player B who engages in his joint project with player A but also betrays her trust.

If B decides to offer the information he sets a price \( p \). Player C observes \( p \), decides whether to buy or not, and if he buys the message and the message turns out to be non-empty he will take a corrective action \( d_C \in \mathbb{R} \) against agent A. For tractability and realism, our model is constructed so that receiving an empty message, or not purchasing it in the first place, is uninformative. This is detailed below.

\subsection{Players’ payoffs as functions of B’s and C’s actions}

In this subsection, we consider the players’ payoffs as they arise from the two economic decisions by players B and C, \( d_B \) and \( d_C \), respectively. We write all three players’ payoffs as depending on some bliss points, which are implicitly defined by the squared payoff reductions. (The parameters \( \kappa, c_A \) and \( c_B \) are further discussed below the formulae.)

\begin{align*}
\pi_A &= -(1 - \kappa)(d_B - \theta_A)^2 - \kappa d_C^2 - c_A(x_A) \\
\pi_B &= -(d_B - \theta_A)^2 + p - c_B \\
\pi_C &= -(d_C - \theta_C \theta_A)^2 - p
\end{align*}

Player A has different bliss points for B’s and C’s actions. She wants B to match her private information \( \theta_A \) and ideally has C choosing 0. This discrepancy generates her main tradeoff and the parameter \( \kappa \in [0, 1] \) governs receive the attention of B. One might think of A’s choice simply as how often A repeats her message (or how fast she sends it), where sending more messages to B will increase the likelihood that he will register them (on time to react). This interpretation relates naturally to the convexity of cost. One might also think of the message being a package of previous communication records, which may or may not contain the interesting bits.
the weights of her two objectives. Players B and C do not care about each other’s actions but each have a bliss point only for their own action. B’s bliss point for his decision is the same as A’s bliss point, $\theta_A$. It is A’s private information and we refer to it as A’s “preference type”. We assume that $\theta_A$’s distribution is strictly increasing, differentiable, and symmetric around zero: $\theta_A \sim F_{\theta_A}$ with existing $f_{\theta_A}(\theta_A) \equiv \frac{dF_{\theta_A}}{d\theta_A}(\theta_A)$ and with $F_{\theta_A}(\theta_A) = 1 - F_{\theta_A}(-\theta_A)$ for all $\theta_A$ in the support which is assumed to be an interval $[-\overline{\theta}_A, \overline{\theta}_A]$. To make the analysis interesting we assume that $\overline{\theta}_A$ is large enough such that at least some types would provide strictly positive amounts of information in the absence of leaks: $\overline{\theta}_A^2(1 - \kappa) > c_A'(0)$.

Player C’s bliss point for his own action is unknown to both players A and B. It is a linear function of A’s preference type, $\theta_C \theta_A$, where $\theta_C$ is C’s private information. That is, C knows how much he cares about $\theta_A$ but he needs knowledge of $\theta_A$ in order to know how to set $d_C$ optimally. We assume that $\theta_C$’s distribution $F_{\theta_C}$ is strictly increasing and differentiable everywhere on its support, which is an interval $[0, \overline{\theta}_C]$. We leave $F_{\theta_C}$ unrestricted except that we impose regularity by assuming that its hazard rate $\frac{f_{\theta_C}}{1-F_{\theta_C}}$ is strictly increasing, where $f_{\theta_C}(\theta_C) \equiv \frac{dF_{\theta_C}}{d\theta_C}(\theta_C)$.

The above assumptions imply that A would like to tell B about $\theta_A$ but prefers C to take minimal actions—which is the situation we had set out to model. Notice that due to the symmetry of $F_{\theta_A}$, the ex-ante optimal action of C is zero—in the absence of obtaining information C will not (and perhaps cannot) take a punitive action. Finally, notice that, as already discussed above, the model is restricted to situations where it is common knowledge that C is not on A’s side. While this is restrictive, it is often the relevant case for issues of privacy.

---

8Letting all bliss points be real-numbered is akin to assuming that the action spaces of B and C are identical. We make this simplifying assumption although it is unlikely to be literally true in many real-world examples. We point out that different action spaces can often be projected onto a one-dimensional scale, which may e.g. measure the preferred degree of job focus.

9Notice that A’s bliss point for C’s action being equal to zero can be assumed without loss of generality: for any other commonly known bliss point $b_{A,C} \in \mathbb{R}$ we can normalize $b_{A,C}$ to be zero if the distribution of is $\theta_A - b_{A,C}$ is symmetric around 0.

10This assumption can also be generated by a more symmetric model where C may be either on A’s side or not, but has a verifiable technology to prove to A if he is on her side.
3.1.2 Information transmission

Both A and B choose whether or not to convey information about $\theta_A$. A chooses whether or not to send B a stochastic message $m_A \in \{\theta_A, \emptyset\}$ that is equal to $\theta_A$ with probability $x_A$ and is empty with probability $1 - x_A$. A’s only choice is the precision $x_A$ of her message. She incurs a cost $c_A(x_A)$ that is increasing and convex and satisfies $c_A(0) = 0$ and $c_A'(0) > 0$.

Player B observes the message $m_A$ and, apart from making his choice of $d_B$, he is also a potential information trader: B decides to offer relaying $m_A$ to C or not, at cost $c_B \in \mathbb{R}^+$. We denote this binary decision by $s_B \in \{0, 1\}$. If he offers $m_A$, i.e. $s_B = 1$, he incurs cost $c_B$ and commits to providing $m_A$ in return for price $p$. $c_B$ reflects the cost of obtaining and supplying data in the market. It is assumed to be irreversible, such that B incurs the cost even if C rejects the offer. We also assume that $p$ and $c_B$ enter the payoffs additively, as described in (2) and (3).

3.1.3 Naivety

As the last ingredient for our model we introduce two different cognitive types of player A, denoted by $\tau_A$, with possible values $n$ (“naive”) and $s$ (“sophisticated”). The naive type occurs with probability $\alpha$ and believes that the cost $c_B$ is higher than it actually is, at $\hat{c}_{B,n} > c_B$. As described in the introduction, information flows from B to C may thus appear to stem from a “leak”. For example, a relevant extreme case is $\hat{c}_{B,n} = \infty$, equivalent to assuming that cognitive type $n$ of Player A ignores the possibility that a market for information exists at all. But overestimation of cost may also appear to a milder extent.

We model this dichotomy of belief types in a game with non-common priors. Although only one cost level is true (and known by players B and C), both cost levels of player B are relevant for all players because each cognitive type of player A believes in one of them. We therefore define two possible cost types of player B, $\tau_B \in \{n, s\}$ with cost levels $c_B \in \{\hat{c}_{B,n}, \hat{c}_{B,s}\}$, and introduce simple (degenerate) non-common prior beliefs where both the sophisticated type of player A as well as player C expect $\hat{c}_{B,s}$ with probability 1 while the naive type of player A expects $\hat{c}_{B,n}$ with probability 1. These beliefs are common knowledge, that is, all higher-order beliefs are “correct”: the different player types agree to disagree in that cognitive type $\tau_A = n$ thinks of $\tau_A = s$ as being paranoid, whereas cognitive type $\tau_A = s$ and players B
and C regard type $\tau_A = n$ as being naive and careless.\footnote{A different modelling approach would have the naive types of player A not understand that other players have different beliefs. For example, naive As might believe that other players believe that all player As are like her. In our model, a naive player A is naive about the costs of information transmission but is sophisticated otherwise. She is not “unaware” like the naive types in the literature on shrouded attributes (Gabaix and Laibson, 2006, Heidhues et al., 2012). We stick to the terminology sophisticated versus naive because in our interpretation of the model she overestimates the true value of $c_B$.} Our welfare analysis will always be based on the true costs, $\hat{c}_{B,s} = c_B$.

Player A’s full type profile is $(\theta_A, \tau_A)$ and we assume that A’s preference type $\theta_A$ is independent of A’s cognitive type $\tau_A$. We also assume that there exist no other stochastic dependence of types between the three players.

## 3.2 Monopoly

We assume that there is only one agent who has access to A’s message and can sell it to the market, i.e., there is only one player B. The sequence of moves in the game is as follows.

- $t = 0$: Each player learns their type.
- $t = 1$: A of type $(\theta_A, \tau_A)$ chooses $x_{A,\tau_A}(\theta_A) \in [0,1]$.
- $t = 2$: Nature chooses the content of $m_A$ and B (of type $\tau_B$) observes it. B chooses $d_B$ and $q_{B,\tau_B} \in [0,1]$, the probability of offering message $m_A$ for sale. If B offers $m_A$, he quotes price $p$ and incurs cost $\hat{c}_{B,\tau_B}$.\footnote{Strictly speaking, we should also index $d_B$ and $p$ with B’s type $\tau_B$. But in equilibrium both types of B choose identical values for these choice variables, hence we drop the indices. As another preview that is relevant for notation, A’s and C’s equilibrium strategy will not involve randomization for almost all types. We therefore introduce mixed-strategy notation only for player B.}
- $t = 3$: If $p$ was quoted by B in $t = 2$, C observes it and chooses $\text{buy}_C(\theta_C) \in \{0,1\}$, indicating whether or not to buy $m_A$. If he buys $m_A$, he observes it and then chooses $d_C(\theta_C)$. All payoffs are realized.

We restrict attention to the game’s symmetric Perfect Bayesian Nash equilibria, where “symmetric” refers to the preference type $\theta_A$; we consider equilibria with $x_{A,\tau_A}(\theta_A) = x_{A,\tau_A}(-\theta_A)$ for all $(\theta_A, \tau_A)$. This restriction to symmetric equilibria implies that neither B nor C adjust $d_B$ and $d_C$ upon observing $m = \emptyset$. (Hereafter, we simply refer to equilibria, dropping “symmetric Perfect Bayesian Nash”.)
An important property of the equilibrium set stems from the observation that there are no equilibria in which B’s profit from selling to C depends on the value of $m_A$. This is because C cannot verify B’s information before making his purchasing decision about $m_A$: if there were equilibria where B’s profit was different for different values of $m_A$ then B could always pretend vis-a-vis C to have the information that induces the highest equilibrium payoff. This property implies that in equilibrium, C does not adjust his beliefs upon observing $p$ and chooses $d_C = 0$ in case he does not buy the message or when it turns out to be empty.

With these observations on the equilibrium set, we can state our first proposition, describing the behavior of C in a monopoly market. All proofs are in the appendix.

**Proposition 1: Equilibrium behavior of player C** There exists a unique $\theta^*_C > 0$ such that if $m_A$ is offered in symmetric equilibrium then player C buys it if and only if $\theta_C \in \left[\theta^*_C, \bar{\theta}_C\right]$.

The proposition is important for our analysis for two reasons, one substantive, one technical. On a substantive level, it shows that A has to brace herself for the worst. The market allocates the information to those types of C whose decision value of the information is highest, i.e. those with the largest equilibrium payoff difference between buying and not buying the information. As the incentives of A and C are opposite, the information will be purchased by those types of C who feel most antagonistic towards A.

The technical relevance lies in the fact that the critical $\theta^*_C$ is independent of other aspects of the equilibrium that is played, i.e. $\theta^*_C$ is constant even across candidate equilibria of different natures. The property is useful for our further analysis because it fixes the expected volume of trade (and the punishment that A will receive from C) conditional on trade occurring. It will translate into the property that the game’s equilibrium is unique.

The proof of Proposition 1 relies on the observation that C’s willingness to pay is given by the function

$$WTP_C(\theta_C, \Gamma_A) = \theta_C^2 \Gamma_A,$$

where the factor $\Gamma_A$ describes the equilibrium value of information and depends only on the collection $\{x_{A,\tau_A}(\theta_A)\}_{\theta_A,\tau_A}$:

$$\Gamma_A \equiv \alpha \int_{\theta_A} x_{A,n}(\theta_A) \theta_A^2 dF_{\theta_A} + (1 - \alpha) \int_{\theta_A} x_{A,s}(\theta_A) \theta_A^2 dF_{\theta_A}$$
With this multiplicative form of $WTP_C$, different candidate equilibrium collections \( \{x_{A,\tau A}(\theta_A)\}_{\theta_A,\tau A} \) merely scale up and down the demand for information that B faces, leaving the critical buyer type unaffected.

Describing the remainder of C’s equilibrium behavior is straightforward: his acquisition induces the choice of his corrective action, \( d_C \), which equals \( \theta_C \theta_A \) if and only if C has acquired \( m_A \) and it is non-empty. For simplicity, we will later notate C’s equilibrium behavior by the phrase “C follows Proposition 1” but mean to imply that his choice \( d_C \) is also optimal in the described way.

Having solved the last part of the game we consider the truncated game played by A and B. This truncated game has the flavor of a simple inspection game. One the one hand, if A expects that B is likely to sell her message, she will be cautious: she withholds information and thereby reduces the market value of information \( \Gamma_A \) to the point where B’s costs are no longer covered such that the market for information may break down. On the other hand, if A expects that there is no market for her information, she will be careless and send her information with maximum precision, raising \( \Gamma_A \) and creating an incentive for C to purchase at prices that make provision worthwhile for B. The equilibrium balances these incentives and, for many parameter constellations, involves mixed strategies. More precisely, the equilibrium predicts the use of mixed strategies for B and a collection of (pure strategies) \( \{x_{A,\tau A}(\theta_A)\}_{\theta_A,\tau A} \) that make B indifferent between selling and not selling.

Technically, we proceed as follows. As \( \theta_C^* \) is constant, each type of player A only requires a belief about the probability with which B sells her message, in order to determine her (A’s) optimal precision. Knowing that B optimally chooses \( d_B = \theta_A \) whenever he observes a nonempty message, and \( d_B = 0 \) otherwise, A’ problem is to maximize

\[
E[\pi_{A,\tau A}(\theta_A)] = -\theta_A^2 (1-\kappa) (1-x_{A,\tau A}) - \theta_A^2 \kappa \hat{q}_{B,\tau A} x_{A,\tau A} \beta_C^2 dF_C(\theta_C) - c_A(x_{A,\tau A})
\]

with respect to \( x_{A,\tau A} \), where \( \hat{q}_{B,\tau A} \) is type \( \tau_A \)’s belief about \( q_B \). Fixing a pair \((q_{B,n}, q_{B,s})\) thus induces the collection \( \{x_{A,\tau A}(\theta_A)\}_{\theta_A,\tau A} \) of best replies to the pair. Conversely, any such collection pins down what the two B types can earn if they sell the information at their given cost and, hence, it pins down their optimal decisions \((q_{B,n}, q_{B,s})\). An equilibrium solves this fixed-point problem from \((q_{B,n}, q_{B,s})\) to \((q_{B,n}, q_{B,s})\). To find it, we consider B’s equilibrium revenue conditional on selling to C:
Definition 1: B’s potential revenue, denoted by $H(q_{B,n}, q_{B,s})$, is the expected monopolistic revenue that B would receive from selling $m_A$ to C, given that simultaneously (i) A believes, and C knows that A believes, that the two types of B offer $m_A$ for sale with probabilities $(\hat{q}_{B,n}, \hat{q}_{B,s}) = (q_{B,n}, q_{B,s})$, (ii) C’s behavior follows Proposition 1, and (iii) A maximizes (6). It is given by

$$H(q_{B,n}, q_{B,s}) = \text{WTP}_C(\theta^*_C, \Gamma_A(q_{B,n}, q_{B,s}))(1 - F_C(\theta^*_C)),$$

where the notation $\Gamma_A(q_{B,n}, q_{B,s})$ indicates that the market value of information depends on $(q_{B,n}, q_{B,s})$.

Notice that $H(q_{B,n}, q_{B,s})$ is identical for both types of B as their costs do not affect their price setting. Notice also that $H(q_{B,n}, q_{B,s})$ does not depend on whether or not B’s actually offers $m_A$ for sale. In other words, it is well-defined only from A’s and C’s behavior. A simple but important implication of (6) is that all types of A will send (weakly) more information as $(q_{B,n}, q_{B,s})$ decreases, which allows establishing that $H$ is decreasing in its arguments. The appendix proves this property and shows how it results in Proposition 2.

Proposition 2: Equilibrium behavior of players A and B
There exists a unique symmetric equilibrium. For no constellation of parameters does the market for information shut player A up: A sends nonempty $m_A$ with strictly positive probability. B’s equilibrium strategy is characterized by the following comparisons, where $0 \leq H(1, 1) < H(0, 1) < H(0, 0)$:

I. If $\hat{c}_{B,n} \leq H(1, 1)$, then $q_{B,n} = q_{B,s} = 1$.

II. If $H(1, 1) < \hat{c}_{B,n} < H(0, 1)$, then $q_{B,n} \in (0, 1)$ and $q_{B,s} = 1$.

III. If $\hat{c}_{B,s} \leq H(0, 1) \leq \hat{c}_{B,n}$, then $q_{B,n} = 0$ and $q_{B,s} = 1$.

IV. If $H(0, 1) < \hat{c}_{B,s} < H(0, 0)$, then $q_{B,n} = 0$ and $q_{B,s} \in (0, 1)$.

V. If $H(0, 0) \leq \hat{c}_{B,s}$, then $q_{B,n} = q_{B,s} = 0$.

Figure 1 illustrates the five regions of cost parameters that lead to the different equilibrium behaviors of B. Notice that differently from Figure 1’s illustration, it may be that $H(1, 1) = 0$, in which case region I fails to exist. This possibility is further discussed below.

14
Figure 1: Equilibrium behavior of player B
We now describe the five regions, one by one. The description also uses some auxiliary results from the appendix: Proposition A.1 shows that a sophisticated \( A \) provides more information than a naive \( A \), i.e. \( x_{A,n}(\theta_A) \geq x_{A,s}(\theta_A) \) for all \( \theta_A \), and that more “surprising” types of \( A \)—those with \( \theta_A \) further away from zero—provide more information, i.e. \( x_{A,\tau_A}(\theta_A) \) increases weakly in \( |\theta_A| \) for all \( \tau_A \). Proposition A.2 describes how the players’ equilibrium payoffs vary with the cost parameters. For the equilibrium description it is also useful to denote by \( x_A(\theta_A,\tau_A) \) and \( \overline{x}_A(\theta_A,\tau_A) \), respectively, the minimal and maximal information provision that player A of type \( (\theta_A,\tau_A) \) chooses in response to the most pessimistic and optimistic expectations about B’s probability of selling to C, \( \hat{q}_{B,\tau_A} = 1 \) and \( \hat{q}_{B,\tau_A} = 0 \). The value of \( x_A(\theta_A,\tau_A) \) may or may not be zero, just as \( \overline{x}_A(\theta_A,\tau_A) \) may or may not be equal to one, depending on the model’s parameters.

I (if it exists): For cost parameters in this region, both types of A anticipate that B sells \( m_A \) for sure. All types of A therefore provide the minimal precision \( x_A(\theta_A,\tau_A) \). The proof of Proposition 2 shows that region I exists iff there exist some types of A whose minimal information provision \( x_A(\theta_A,\tau_A) \) is strictly positive, and that this happens iff it is true for the most extreme preference type, \( \bar{\theta}_A \). More precisely, the region exists if

\[
1 - \kappa(1 + \int_{\theta_C}^{\bar{\theta}_C} \theta_C^2 dF_C(\theta_C)) \geq \frac{c'(0)}{\bar{\theta}_A^2} \tag{7}
\]

and otherwise, only regions II to V exist.

II: In this parameter region, the sophisticated A continues to correctly anticipate that B sells for sure and hence chooses \( x_A(\theta_A,s) \). But the naive A is more optimistic (and less careful) about the behavior of his imaginary opponent, high-cost B. The condition \( H(1,1) < c_{B,n} \) implies that \( \hat{q}_{B,n} < 1 \) and every \( \theta_A \)-type of the naive A plays her (pure-strategy) best response to \( \hat{q}_{B,n} \), which lies strictly above \( \overline{x}_A(\theta_A,n) \) for some \( \theta_A \). The high-cost B’s cost, \( \hat{c}_{B,n} \), is larger than in region I and he therefore has an incentive to reduce \( q_{B,n} \) as long as \( H(q_{B,n},1) < c_{B,n} \). In equilibrium, he makes himself indifferent by setting \( q_{B,n} \) such that \( H(q_{B,n},1) = c_{B,n} \). The (true) low-cost B’s profits in region II exceed those of region I as the additional information that the naive A provides creates additional demand by C.

\(^{13}(6)\) implies that only a zero-measure set of naive A types are indifferent between different precision levels.
III: In this region the naive type of player A is even more optimistic about the cost of information transmission. She feels uninhibited in providing information, \( x_{A,n}(\theta_A) = \pi_A(\theta_A, n) \). The sophisticated A anticipates that B sells the information for sure and chooses the minimal information provision \( x_A(\theta_A, s) \). Player B sells the minimal information provided by sophisticated As, plus the maximal information provided by the naive As. In terms of profits this is B’s best parameter region.

IV: While naive A still chooses maximal \( x_A(\theta_A, n) \) (as she assume that B will never sell), now the sophisticated A plays a game of cat and mouse with low-cost B. The logic for the interaction between the sophisticated A and (real) low-cost B is analogous to the mixed-equilibrium logic of region II: sophisticated A provides just enough information to make B indifferent between selling and not, and B sells with low enough probability to not reduce the information provision too far. But there is an important difference regarding the actual profits of (low-cost) B. In the game of cat and mouse in region IV, the sophisticated A correctly considers the revenue that B can generate from selling information sent by the naive A. Consequently, in expectation, the low-cost B makes zero profits in this equilibrium.

V: While desirable from the perspective of As, this is the least interesting equilibrium region. B’s true costs are prohibitive and so are the even higher costs imagined by the naive A player. There is no market for information. All types of A choose maximal information precision \( \pi_A(\theta_A, n) \) and achieve their highest possible payoff, while player B earns nothing as a trader of information.

We now consider a selection of comparative statics results. As indicated above, the appendix contains a comprehensive result on comparative statics, Proposition A.2, regarding the players’ payoffs when moving within and between the different regions. In reality, such moves may be generated by regulatory intervention that affect the actual cost of B entering the market for information, \( \hat{c}_{B,s} = c_B \), and/or the perceived cost \( \hat{c}_{B,n} \). The following corollary considers the consequences of increasing both \( \hat{c}_{B,s} \) and \( \hat{c}_{B,n} \) by constant proportions, that is, along some ray \( \hat{c}_{B,n} = \lambda \hat{c}_{B,s} \) with \( \lambda > 1 \). In Figure 1, this corresponds to moving northeast on a straight line that starts at the origin and lies above the 45-degree line.

**Corollary 1** For sufficiently low \((\hat{c}_{B,n}, \hat{c}_{B,s})\),

(i) a modest proportional increase in \((\hat{c}_{B,n}, \hat{c}_{B,s})\) weakly reduces naive A’s
payoff, for all $\theta_A$, without increasing sophisticated A’s payoff;

(ii) a sufficiently large proportional increase in $(\hat{c}_{B,n}, \hat{c}_{B,s})$ weakly increases
the payoffs of all types of player A.

Clause (i) demonstrates that a half-baked regulatory invention would only
harm the naive player A, as she would feel safe too soon. The result derives
from the observation that a modest proportional increase in $(c_{B,n}, c_{B,s})$ corre-
sponds to a northeast move within region II or a move from region I to region
II. The cost level weakly increases the potential equilibrium revenue $H$ be-
cause high-cost B has to be indifferent in region II. By weak monotonicity
of $H$, this means that $q_{B,n}$ is reduced, inducing a weakly higher $x_{A,n}$. Since
the true probability of selling is $q_{B,s} = 1$, this leads to a payoff reduction
(otherwise, sophisticated A would increase $x_{A,s}$, too. Clause (ii) refers to a
strong measure of privacy protection, moving from region I or II to region V,
where all types of A have maximal payoffs.

Somewhat counter-intuitively, it may not be possible to induce Pareto im-
provement through an increase in $(\hat{c}_{B,n}, \hat{c}_{B,s})$, relative to a situation where A’s
personal data is relatively unprotected. One may have thought that stronger
data protection enables the players to move from a completely silent equilib-
rrium to one where an intermediate level of information is shared and traded,
to everyone’s benefit. But as Proposition 2 shows, the equilibrium never al-

tows for complete silence: in equilibrium regions I and II, players B and C
benefit from the naive players who chatter away. Moving towards regions IV
and V would therefore not necessarily improve B’s and C’ payoffs.\footnote{This
does not mean that no Pareto improvement is possible, but we believe that one
would need stronger assumptions to show its existence.}

Our next corollary concerns variations in the share of naive players, which
might be affected by policy interventions such as educational campaigns
about the actual costs of surveillance and data provision.

**Corollary 2** An increase in $\alpha$

(i) reduces information provision by any given type:

\[
\frac{dx_{A,n}(\theta_A)}{d\alpha} \leq 0 \quad \text{and} \quad \frac{dx_{A,s}(\theta_A)}{d\alpha} \leq 0, \forall \theta_A
\]

(ii) harms the sophisticated A: \( \frac{d\pi_{A,s}(\theta_A)}{d\alpha} \leq 0, \forall \theta_A \)
(iii) can benefit the naive: \( \exists \alpha', \alpha \text{ such that } \alpha' > \alpha \) and \( \pi_{A,n}(\theta_A) \) is strictly larger at \( \alpha' \) than at \( \alpha \), for all \( \theta_A \).

The corollary shows how all types of A compensate for a large prevalence of naivete. A larger share of naive players wets the appetite of C to buy the available information, and hence the likelihood that B is willing to sell. As a result, both naive and sophisticated A reduce their information provision (clause (i)). The overall value of the available information \( \Gamma_A \), however, increases in response to a larger \( \alpha \) because naive agents provide more information in the aggregate. (Otherwise, if \( \Gamma_A \) was reduced, then all types of A would increase their \( x_A \). Since naive A provides more information than sophisticated A, and naive A’s share increases, the resulting aggregate level of \( \Gamma_A \) would have to increase, a contradiction.) For the sophisticated A, this effect translates into a reduction of the equilibrium payoff, in response to larger \( \alpha \). In contrast, the increase in \( \alpha \) may help the naive players. This is possible because increase in \( \alpha \) affect the boundaries of the equilibrium regions and the increase in \( \alpha \) may thus corresponds to a change from the equilibrium prediction in region I to the equilibrium prediction in region II: the increase in \( \alpha \) makes naive player A behave more cautiously and reduce their loss.

Our final, and brief, comparative statics exercise examines a variation in the degree to which player C cares about A’s information. What if C becomes more or less aggressive towards A? That is, we consider the consequences of scaling the distribution of \( \theta_C \). Let \( \theta_C = \mu \theta_C \) and \( \tilde{F}(\theta_C) = F(\theta_C) \). Player C’s willingness to pay for the information now becomes \( WTP(\tilde{\theta}_C) = \mu^2 \theta_C^2 \Gamma_A \). Scaling the distribution of C types is therefore equivalent to shifting \( \Gamma_A \) and leaves the selection of buyers in B’s monopoly untouched: for the distribution \( \tilde{F} \) the marginal type to which B sells is simply given by \( \tilde{\theta}_C^* = \mu \theta_C^* \). The consequences for A’s choice of information precision are, hence, straightforward. For \( \mu > 1 \) (\( \mu < 1 \)) her information becomes more (less) valuable in the market and the potential punishment more (less) severe. Thus, she will weakly reduce (increase) her precision and will be weakly worse (better) off.

4 Conclusion

The ongoing public discussions about trading of personalized information include many fascinating examples. A recent item concerns the data provision
by mobile phone applications to data aggregators. It is a typical example of an ill-understood market mechanism. Very few telephone users ask themselves why free-of-charge software is provided in their Appstore. The answer they would get (if they cared to ask) is that these apps routinely transfer personalized data from the telephone to data aggregators, usually for advertisement purposes, and usually in return for money.

The business model 'service-in-exchange-for-data' is not new, and we use it here as a particular interpretation of our model: B may be an app provider, bringing a service to the telephone user but selling his information on to C. The interpretation illustrates that data aggregators need to offer the users a benefit, like a useful software, and that this necessity to engage the customer disciplines the industry. Without receiving a benefit, few rational customers would provide the data. But anyway, few people are rational in their app usage: naive users do not notice that they constantly provide information to the hidden data aggregator. As our analysis indicates, it is likely that the industry’s larger profits lie in dealing with naive customers who do not withhold information. That is, the data trading industry has strong incentives to create markets where selling information is (still) unanticipated by the originators of information. In markets for new IT products, this is typically the case. But even sophisticated customers suffer because of the externality that we discuss in this paper: the demand for data interception services increases with the degree of naivete in the population.
References


A Appendix A: Proofs

Proof of Proposition 1

Consider C’s choice of $d_C$. The uncertainty that he faces stems from the multidimensional randomness in $x_A$, $\theta_A$ and $\tau_A$. For a notation that avoids multiple integrals, we denote by $\xi$ a random variable that governs all of this randomness and we let $\mathfrak{B}_\xi$ be its support and $F_\xi$ its distribution function.

C’s optimal strategy $d_C^*$ is a function of $(\xi, \theta_C)$. As indicated in the main text, $p$ does not carry any relevant information and the symmetry of $x_A, \tau_A(\theta_A)$ around $\theta_A = 0$ implies that C’s optimal action is $d_C^* = 0$ whenever he does not buy $m_A$ or if it is empty. Let $\mathfrak{B}_\xi^0$ be the subset of $\mathfrak{B}_\xi$ where $d_C^* = 0$ if C buys $m_A$. The “other” values of $\xi$, $\xi \notin \mathfrak{B}_\xi^0$, are those where C, if he buys $m_A$, observes a nonempty and nonzero value $m_A = \theta_A \neq 0$, and optimally sets $d^*_C(\theta_C) = \theta_C \theta_A$. We can then write C’s expected payoff from buying $m_A$ and responding to it with his optimal strategy $d_C^*$ as

$$E[\pi_C\{buy_C = 1, d_C^*, \theta_C\}] = \int_{\xi \in \mathfrak{B}_\xi} -(d_C^* - \theta_C \theta_A)^2 dF_\xi - p$$

$$= \int_{\xi \in \mathfrak{B}_\xi^0} -(0 - \theta_C \theta_A)^2 dF_\xi + \int_{\xi \notin \mathfrak{B}_\xi^0} -(\theta_C \theta_A - \theta_C \theta_A)^2 dF_\xi - p$$

$$= \int_{\xi \in \mathfrak{B}_\xi^0} -(\theta_C \theta_A)^2 dF_\xi - p$$
C’s payoff from not buying $m_A$ is

$$E[\pi_C(\text{buy}_C = 0, d^*_C, \theta_C)] = \int_{\xi \in \mathbb{B}_\xi} - (\theta_C \theta_A)^2 dF_\xi$$

and his willingness to pay for $m_A$ is therefore:

$$WTP_C(\theta_C) = E[\pi_C(\text{buy}_C = 1, d^*_C, \theta_C)] - E[\pi_C(\text{buy}_C = 0, d^*_C, \theta_C)]$$

$$= \int_{\xi \in \mathbb{B}_\xi} - (\theta_C \theta_A)^2 dF_\xi - p + \int_{\xi \in \mathbb{B}_\xi} - (\theta_C \theta_A)^2 dF_\xi$$

$$= \theta_C^2 \int_{\xi \notin \mathbb{B}_\xi} \theta_A^2 dF_\xi$$

Now we give up the $\xi$-Notation and ask about the random outcomes that correspond to $\xi \notin \mathbb{B}_\xi$. When is $m_A$ nonempty? All of the randomness in $x_A$, $\theta_A$ and $\tau_A$ is independent by assumption. Aggregating across $\tau_A$ yields the $\theta_A$-specific precision of $m_A$. Aggregating further across $\theta_A$ gives $m_A$’s overall average probability of being nonempty as

$$\int_{\xi \notin \mathbb{B}_\xi} dF_\xi = \alpha \int_{\theta_A} x_{A,n}(\theta_A) dF_{\theta_A} + (1 - \alpha) \int_{\theta_A} x_{A,s}(\theta_A) dF_{\theta_A}$$

C’s willingness to pay is analogous to this expression but weights each $\xi \notin \mathbb{B}_\xi$ by $\theta_A^2$, and weights the resulting integral by $\theta_C^2$. That is, we can rewrite C’s willingness to pay as

$$WTP_C(\theta_C, \Gamma_A) = \theta_C^2 \Gamma_A$$

where

$$\Gamma_A \equiv \alpha \int_{\theta_A} x_{A,n}(\theta_A) \theta_A^2 dF_{\theta_A} + (1 - \alpha) \int_{\theta_A} x_{A,s}(\theta_A) \theta_A^2 dF_{\theta_A}$$

(which is (5)). The function $WTP_C(\theta_C, \Gamma_A)$ is strictly increasing in $\theta_C$, in any equilibrium with $\Gamma_A > 0$. Therefore, if $\Gamma_A > 0$, there exists a unique threshold

$$\theta_C^*(\Gamma_A, p) \equiv \sqrt{\frac{p}{\Gamma_A}} > 0$$

such that C buys $m_A$ if and only if $\theta_C \geq \theta_C^*(\Gamma_A, p)$. If $\Gamma_A = 0$, C’s willingness to pay is zero and B would therefore not offer $m_A$.
We now show that \( \theta^*_C(\Gamma_A, p) \) is independent of \( \Gamma_A > 0 \) whenever \( p \) is the monopoly price. This will complete the proof because in equilibrium, \( p \) is the monopoly price and \( \Gamma_A > 0 \) must hold if \( B \) is willing to offer \( m_A \) (and incur cost \( c_B, \tau_B \)).

For given \( \Gamma_A \), consider \( WTP_C(\theta_C, \Gamma_A) \) as a random variable (driven by randomness in \( \theta_C \)). Denoting its distribution by \( F^\Gamma_{WTP_C} \) and the latter’s derivative by \( f^\Gamma_{WTP_C} \), we first argue that \( WTP_C \) has a monotone hazard rate. Observe that by definition it holds for all \( \theta_C \):

\[
F^\Gamma_{WTP_C}(WTP_C(\theta_C, \Gamma_A)) = F_{\theta_C}(\theta_C)
\]

Differentiating both sides with respect to \( \theta_C \), we obtain \( f^\Gamma_{WTP_C} \):

\[
f^\Gamma_{WTP_C}(WTP_C(\theta_C, \Gamma_A)) \frac{\partial WTP_C(\theta_C, \Gamma_A)}{\partial \theta_C} = f_{\theta_C}(\theta_C)
\]

\[
f^\Gamma_{WTP_C}(WTP_C(\theta_C, \Gamma_A)) = f_{\theta_C}(\theta_C) \frac{\partial \theta_C}{\partial WTP_C(\theta_C, \Gamma_A)}
\]

For an increasing hazard rate of \( WTP_C \), we therefore require:

\[
\frac{\partial}{\partial WTP_C} \frac{f^\Gamma_{WTP_C}(WTP_C(\theta_C, \Gamma_A))}{1 - F^\Gamma_{WTP_C}(WTP_C(\theta_C, \Gamma_A))} > 0
\]

\[
\frac{\partial}{\partial WTP_C} \left[ \frac{f_{\theta_C}(\theta_C)}{1 - F_{\theta_C}(\theta_C)} \frac{\partial \theta_C}{\partial WTP_C(\theta_C, \Gamma_A)} \right] > 0
\]

For a differentiable function \( h(\theta_C) \) we have that \( \frac{\partial}{\partial WTP_C}[h] = \frac{\partial}{\partial \theta_C}[h \frac{\partial \theta_C}{\partial WTP_C}] \) and thus we have an increasing hazard rate of \( WTP_C \) if:

\[
\frac{\partial}{\partial \theta_C} \left[ \frac{f_{\theta_C}(\theta_C)}{1 - F_{\theta_C}(\theta_C)} \left( \frac{\partial \theta_C}{\partial WTP_C(\theta_C, \Gamma_A)} \right)^2 \right] > 0
\]

\[
\frac{\partial}{\partial \theta_C} \frac{f_{\theta_C}(\theta_C)}{1 - F_{\theta_C}(\theta_C)} > 0,
\]

This is true by assumption of \( \theta_C \)'s strictly increasing hazard rate and thus \( WTP_C \), too, has a strictly increasing hazard rate.

We now show that \( \theta^*_C \) is identical for different values of \( \Gamma_A \) that may arise in equilibrium. Towards a contradiction, suppose that there exist two equilibria of the game (equil. (1)) and (equil. (2)) that result in values
\[(\Gamma_A^{(1)}, \Gamma_A^{(2)})\] with corresponding optimal monopoly prices \((p^{(1)}, p^{(2)})\) and with marginal buyers \((\theta_C^{(1)} = \theta_C^{*}(\Gamma_A^{(1)}, p^{(1)}), \theta_C^{(2)} = \theta_C^{*}(\Gamma_A^{(2)}, p^{(2)})\) such that \(\theta_C^{*}(1) > \theta_C^{*}(2)\). That is, we have the following indifference conditions:

\[\begin{align*}
p^{(1)} &= (\theta_C^{*}(1))^2 \Gamma_A^{(1)} \\
p^{(2)} &= (\theta_C^{*}(2))^2 \Gamma_A^{(2)}
\end{align*}\]  

Since \(WTP_C\) has a strictly increasing hazard rate, we have that \(p^{(2)}\) is uniquely optimal for \(\Gamma_A^{(2)}\) (see e.g. Börgers, 2015). Thus, in equil. (2) it is strictly better for B to make \(\theta_C^{*}(2)\) indifferent than to make \(\theta_C^{*}(1)\) indifferent. To achieve the latter, B would have to set a price satisfying

\[
\tilde{p} = (\theta_C^{*}(1))^2 \Gamma_A^{(2)} = \frac{\Gamma_A^{(2)}}{\Gamma_A^{(1)}} p^{(1)}
\]

where the latter equality follows from plugging in (8). The assumption that \(p^{(2)}\) is strictly better than \(\tilde{p}\) in equil. (2) applies therefore implies the following revenue inequality:

\[p^{(2)}(1 - F_{\text{WTP}}^{(1)}(WTP_C(\theta_C^{*}(1), \Gamma_A^{(2)}))) > \tilde{p}(1 - F_{\text{WTP}}^{(2)}(WTP_C(\theta_C^{*}(1), \Gamma_A^{(2)})))(10)\]

\[p^{(2)}(1 - F_{\theta C}(\theta_C^{*}(2))) > \tilde{p}(1 - F_{\theta C}(\theta_C^{*}(1)))(11)\]

\[
\frac{\Gamma_A^{(1)}}{\Gamma_A^{(2)}} p^{(2)} (1 - F_{\text{WTP}}^{(1)}(WTP_C(\theta_C^{*}(1), \Gamma_A^{(2)}))) > p^{(1)} (1 - F_{\text{WTP}}^{(1)}(WTP_C(\theta_C^{*}(1), \Gamma_A^{(2)})))(12)
\]

Now observe that that the LHS of (13) is a feasible revenue in B’s maximization problem in equil. (1): multiply both sides of (9) by \(\frac{\Gamma_A^{(1)}}{\Gamma_A^{(2)}}\) to see that \(\frac{\Gamma_A^{(1)}}{\Gamma_A^{(2)}} p^{(2)}\) is the price that makes \(\theta_C^{*}(2)\) indifferent in equil. (1):

\[
\frac{\Gamma_A^{(1)}}{\Gamma_A^{(2)}} p^{(2)} = (\theta_C^{*}(2))^2 \Gamma_A^{(1)}
\]

All types higher than \(\theta_C^{*}(2)\) will therefore also buy and hence the LHS of (13) is the revenue that B gets from setting price \(\frac{\Gamma_A^{(1)}}{\Gamma_A^{(2)}} p^{(2)}\). This implies that the
RHS of (13) is not the optimal revenue for $\Gamma_A^{(1)}$, contradicting our initial assumption of $p^{(1)}$ being the monopoly price. ■

Proof of Proposition 2

Like in the main text, we focus on the truncated game played by $A$ and $B$ after having described $C$’s behavior in Proposition 1. The justification for this simplification is recaptured by the observations in this paragraph. In the monopoly game between $B$ and $C$, they both take $A$’s equilibrium behavior as given, which results in a value of information $\Gamma_A$, according to (5). Both types of $B$ optimally set the same price $WTP_C(\theta^*, \Gamma_A)$ if they decide to sell the information, and $C$ follows Proposition 1. As is explained in the main text, the price does not contain information about the message $m_A$ so $C$’s beliefs, along the equilibrium path, will not be updated upon observing the price. We also have to briefly consider off-equilibrium beliefs. They are unproblematic because the equilibrium must prescribe that if $C$ observes a price that should not arise in equilibrium, he sticks to the strategy to buy if and only if $\theta_C \geq \theta_C^*(\Gamma_A, p)$. This is the uniquely sequentially rational strategy, as otherwise $C$ could be tricked into buying packages that are too expensive for him.

Now fix beliefs of player $A$ about the likelihood of $B$ selling message $m_A$ to $C$. In equilibrium, the two different cognitive types $\tau_A$ have different beliefs, but different $\theta_A$ do not. Hence, denote beliefs of type $\tau_A$ by $\hat{q}_{B,\tau_A}$ and observe that she maximizes (6),

$$E[\pi_{A,\tau_A}(\theta_A)] = -\theta_A^2(1-\kappa)(1-x_{A,\tau_A})-\theta_A^2\kappa \hat{q}_{B,\tau_A}x_{A,\tau_A} \int_{\theta_C^*}^{\tilde{\theta}_C} \theta_C^2 dFC(\theta_C) - c_A(x_{A,\tau_A}),$$

with respect to $x_{A,\tau_A}$. The first two terms of this objective are linear in $x_{A,\tau_A}$ and the third term is twice continuously differentiable and concave in $x_{A,\tau_A}$. Hence, $E[\pi_{A,\tau_A}(\theta_A)]$ has a unique maximizer $x_{A,\tau_A}(\theta_A)$. Moreover, at an interior solution the maximizer is continuous and differentiable in $q_{B,\tau_A}$, weakly decreases in $q_{B,\tau_A}$ and strictly decreases in $q_{B,\tau_A}$.

Next consider $B$’s problem. The two types of $B$ simply compare their expected revenue from selling the message, $H(q_{B,n}, q_{B,s})$ to their cost. If

$$H(q_{B,n}, q_{B,s}) \geq \tilde{c}_{B,\tau_B} \quad (14)$$

26
holds strictly he chooses \( q_{B,\tau_B} = 1 \); if it does not hold he chooses \( q_{B,\tau_B} = 0 \) and if it holds with equality he can mix. Notice that the latter cannot simultaneously be the case for both levels of \( \tau_B \) because \( \tau_B = n \) has a strictly higher cost than \( \tau_B = s \). If the low-cost type mixes, the high-cost type does not sell the message. Conversely, if the high-cost type mixes, the low-cost type sells it for sure. It follows that there exist only the five different equilibrium configurations of \((q_{B,n}, q_{B,s})\) that the proposition lists. It remains to show that the equilibrium is unique and that condition (7) determines whether all five, or only four, equilibrium configurations can arise for possible levels of \( \hat{c}_{B,\tau_A} \).

To establish uniqueness, we first notice that \( H(q_{B,n}, q_{B,s}) \) strictly decreases in both arguments if they lie in the open interval \((q, \bar{q})\) where \( q \) is the largest value of \( \hat{q}_{B,\tau_A} \) that induces maximal information precision for all types of \( A \) and \( \bar{q} \) is the smallest value of \( \hat{q}_{B,\tau_A} \) that induces minimal information precision. This strict monotonicity follows from the above observation that the optimal \( x_{A,\tau_A}(\theta_A) \) weakly decreases in \( \hat{q}_{B,\tau_A} \) and strictly decreases at an interior solution, for all \((\theta_A, \tau_A)\): an increase in \( \hat{q}_{B,\tau_A} \) induces all \( A \) types to be more cautious and it induces all \( A \) types who provide some information to be strictly more cautious. Since the revenue is \( H = WTP_C(\theta^*, \Gamma_A)(1 - F_{\theta_C}(\theta^*)) \), it follows from (4) and (5) that \( H \) decreases strictly in \((q_{B,n}, q_{B,s})\) if all elements of the collection \( \{x_{A,\tau_A}(\theta_A)\}_{\theta_A,\tau_A} \) decrease weakly in \((q_{B,n}, q_{B,s})\) and some of them strictly.

We now show that \( q = 0 \). For any possible level of information provision \( \bar{x}_A \geq 0 \), the strict concavity of (6) implies that information provision exceeds \( \bar{x}_A \) at belief \( \hat{q}_{B,\tau_A} \) iff:

\[
\frac{\partial}{\partial x_A} [-\theta_A^2(1 - \kappa)(1 - x_{A,\tau_A}) - \theta_A^2 \kappa \hat{q}_{B,\tau_A} x_{A,\tau_A} \int_{\theta_C}^{\bar{\theta}_C} \tilde{\theta}_C^2 dF_C(\theta_C)] \geq c'_A(\bar{x}_A)
\]

\[
\theta_A^2(1 - \kappa) - \theta_A^2 \kappa \hat{q}_{B,\tau_A} \int_{\theta_C}^{\bar{\theta}_C} \tilde{\theta}_C^2 dF_C(\theta_C) \geq c'_A(\bar{x}_A)
\]

But for \( \theta_A \) close to zero, the LHS is close to zero independently of \( \hat{q}_{B,\tau_A} \) and \( \bar{x}_A \), which, together with \( c'(\cdot) > 0 \) implies that the inequality cannot hold for all \( \theta_A \). To induce the largest possible set of \( \theta_A \) types to provide \( \bar{x}_A \), we need to reduce \( \hat{q}_{B,\tau_A} \) and \( \bar{x}_A \) to their minimal levels, which are zero in both cases. That is, \( q = 0 \) and the marginal type \( \theta_A \) who provides strictly positive
information at \( q \) satisfies \( \theta_A^2(1 - \kappa) = c'_A(0) \). A type with a large enough \( \theta_A \) to satisfy this equality exists by assumption.

The upper limit of the named interval, \( \bar{q} \), may or may not lie strictly below 1. The case \( \bar{q} < 1 \) occurs if the most extreme preference type \( \bar{\theta}_A \) does not provide any information at belief \( \hat{q}_{B,\tau_A} = 1 \) i.e. iff

\[
\frac{\partial}{\partial x_A}[-\bar{\theta}_A^2(1 - \kappa)(1 - x_{A,\tau}) - \bar{\theta}_A^2 \kappa x_{A,\tau} A \int_{\theta_C} \theta_C dF_C(\theta_C)] < c'(0)
\]

which is the negation of condition (7). If \( \bar{q} < 1 \), then by definition of \( \bar{q} \) no type of A reduces \( x_A \) any further as \( \hat{q}_{B,\tau_A} \) rises above \( \bar{q} \). This implies, using (6) and the fact that \( c_A \) is twice continuously differentiable, that the optimal \( x_{A,\tau}(\theta_A) \) cannot be at an interior optimum if \( \hat{q}_{B,\tau_A} > \bar{q} \). This implies that \( x_{A,\tau}(\theta_A) = 0 \) holds for all \( \theta_A \) if \( 1 \geq \hat{q}_{B,\tau_A} > \bar{q} \). In this case, C's (equilibrium) willingness to pay for messages sent by A players of cognitive type \( \tau_A \) is zero, too, and thus \( H \) is constant in \( \hat{q}_{B,\tau_A} \) at \( \hat{q}_{B,\tau_A} > \bar{q} \). Moreover, if \( 1 \geq \hat{q}_{B,\tau_A} > \bar{q} \) holds for both cognitive types \( \tau_A \), then \( H(q_{B,n}, q_{B,s}) = 0 \).

Summing up the previous two paragraphs, we see that \( H(q_{B,n}, q_{B,s}) \) strictly decreases in both arguments if they lie in \([0, \bar{q}]\). \( \bar{q} \) may lie strictly below 1, which happens if (7) does not hold. In this case \( H(q_{B,n}, q_{B,s}) \) is constant in both arguments if they lie above \( \bar{q} \), and \( H(q_{B,n}, q_{B,s}) = 0 \) if both arguments lie above \( \bar{q} \).

The monotonicity of \( H \) implies that the equilibrium is unique. Intuitively, if (14) has slack, then B of type \( \tau_B \) increases \( q_{B,\tau_B} \), and the monotonicity of \( H(q_{B,n}, q_{B,s}) \) ensures that (14) holds with equality at no more than one combination \( (q_{B,n}, q_{B,s}) \). Towards a contradiction, suppose that there exist two equilibria \( (q_{B,n}^{(1)}, q_{B,s}^{(1)}), (q_{B,n}^{(2)}, q_{B,s}^{(2)}) \). Different probabilities of \( \tau_B \) selling can only occur if \( H(q_{B,n}, q_{B,s}) = \hat{c}_{B,\tau_B} \) holds in both equilibria. Therefore, both equilibria must yield the same revenue \( H \). With the above monotonicity properties of \( H \), this can arise under exactly two circumstances: either all four probabilities \( (q_{B,n}^{(i)}, q_{B,s}^{(i)}, q_{B,n}^{(j)}, q_{B,s}^{(j)}) \) lie above \( \bar{q} \), or \( q_{B,n}^{(i)} < q_{B,n}^{(j)} \) and \( q_{B,s}^{(i)} > q_{B,s}^{(j)} \) holds for \( i, j \in \{1, 2\} \) and \( j \neq i \). In the former case, revenue \( H \) is zero and
hence all four probabilities must be zero and the two equilibria cannot differ. The latter case implies that both types are indifferent in equilibrium, which is impossible because $\hat{c}_{B,n} > \hat{c}_{B,s}$. This establishes uniqueness.

The borders of the equilibrium regions follow trivially from condition (14). The critical value $H(1, 1)$ is strictly positive iff at least some types of A provide strictly positive information if $m_A$ is sold for sure, i.e. if (7) holds. In this case, it can be that $H(1, 1) \geq \hat{c}_{B,n}$ and the equilibrium is of type I. Regions II, III, IV and V exist for sure (because $H(0, 0)$ and $H(0, 1)$ are strictly positive) and they, respectively, are relevant and correspond to B’s behavior as specified in Proposition 2 as follows. II: $H(0, 1) > \hat{c}_{B,n} > H(1, 1)$. III: $\hat{c}_{B,s} \leq H(0, 1)$ and $\hat{c}_{B,n} \geq H(0, 1)$. IV: $H(0, 0) > \hat{c}_{B,s} > H(0, 1)$. V: $H(0, 0) \leq \hat{c}_{B,s}$. ■

**Proposition A.1** In any symmetric equilibrium:

(i) Naive A types provide more: $x_{A,n}(\theta_A) \geq x_{A,s}(\theta_A), \forall \theta_A$

(ii) Interesting A types provide more: $x_{A,\tau_A}(\theta_A)$ increases weakly in $|\theta_A|, \forall \tau_A$

**Proof** Both results are almost immediate from A’s objective function in (6). Clause (i) uses the observation that $\hat{q}_{B,n} \geq \hat{q}_{B,s}$ (see Proof of Proposition 2), which implies the result together with (6). Clause (ii) follows from (6) by observing that A’s benefit from transmitting information is linear in $x_A$ and quadratic in $\theta_A$: the optimal information provision is 0 if

$$1 - \kappa(1 + \hat{q}_{B,\tau_A}) \int_{\theta_{C}^*}^{\tilde{\theta}_{C}} \theta_{C}^2 dF_C(\theta_C) < \frac{c'(0)}{\theta_{A}^2},$$

and otherwise it increases strictly in $\theta_{A}^2$. ■

We now add a proposition about the players’ (expected) equilibrium payoffs as we vary $(\hat{c}_{B,n}, \hat{c}_{B,s})$. (We drop arguments and expectation operators for brevity.) In order to accommodate the two interpretations for player B (see the introduction), we split his payoff into two parts, $\pi_{B_1}$ resulting from the direct interaction with player A and $\pi_{B_2}$ resulting from the sale of the message to player C.

**Proposition A.2**
Proof For the proof of the proposition, we first introduce two Lemmas.

Lemma 1 \( \frac{d\pi_{A,s}}{dc_{B,s}} \geq 0 \) and \( \frac{d\pi_{A,n}}{dc_{B,n}} = 0 \)

Proof of Lemma 1 Let us first examine the first part of the lemma. Take \( \tau_B = n \). For \( c_{B,s} \geq H(0,1) \) the high-cost B never sells the message. For \( c_{B,s} < H(0,1) \) we know that as we increase \( c_{B,n} \) from zero upwards we are first in equilibrium region I where the high-cost B always sells, then in region II where he mixes, and finally in region III where he never sells. Hence, the proof will be complete if we can show that the claim holds within region II. In this region, the high-cost B is indifferent, i.e. \( H(q_{B,n}, q_{B,s}) = c_{B,n} \). Recall from the Proof of Proposition 2 that \( H \) is monotonically decreasing in both arguments, which yields the result for \( \tau_B = n \). A similar logic applies to \( \tau_B = s \).

The next lemma shows that for an arbitrary change in parameters, the payoff of sophisticated A and the profits that B obtains from his direct interaction with A move in parallel with (i.e. their difference has the same sign as the difference in) sophisticated A’s information precision

Lemma 2 For all \( \theta_A \) and for any parameter \( \gamma \) that is either in \( \{ \kappa, \alpha, \hat{c}_{B,n}, \hat{c}_{B,s} \} \) or is a distribution parameter of \( F_A \) or \( F_C \) (but not a parameter of A’s cost function): \( \text{sign}(d\pi_{A,s}(\theta_A)/d\gamma) = \text{sign}(d\pi_{B,s}/d\gamma) = \text{sign}(dx_{A,s}(\theta_A)/d\gamma) \).
Proof of Lemma 2 To see that $\text{sign}(d\pi_{B_1}/d\gamma) = \text{sign}(dx_{A,s}(\theta_A)/d\gamma)$, observe from (6) that the information provision of sophisticated A moves in the same direction for all $\theta_A$, i.e. $\text{sign}(dx_{A,s}(\theta_A)/d\gamma) = \text{sign}(dx_{A,s}(\theta'_A)/d\gamma)$ for all $(\theta_A, \theta'_A)$. Therefore, B benefits more from his interaction with sophisticated A iff $dx_{A,s}(\theta_A)/d\gamma \geq 0$. For the result to be wrong, we would need that naive A strictly adjusts her information provision in the opposite direction. But in equilibrium, the beliefs about B need to describe B’s selling behavior as optimal according to (14). There cannot be a change in $H$ or in one of the cost parameters $(\hat{c}_{B,n}, \hat{c}_{B,s})$ that would strictly add slack to (14) for one type of B and strictly reduce slack for the other type. Hence (using (6) again) there cannot exist a change in $\gamma$ that induces a strict increase in $x_{A,s}$ but a strict reduction in $x_{A,n}$, or vice versa.

To see that $\text{sign}(d\pi_{A,s}(\theta_A)/d\gamma) = \text{sign}(dx_{A,s}(\theta_A)/d\gamma)$, consider sophisticated A’s expected payoff in (6). Its first two parts are linear in $x_{A,s}$ and the convex cost of $x_{A,s}$ is subtracted. Hence, A’s payoff equals the area between a constant line (reflecting the marginal benefits of transmitting information) and some increasing function capturing A’s marginal costs. The latter remains unaffected by $\gamma$. So any change in $\gamma$ will only move the constant marginal benefit. (6) implies that if $\gamma$ moves that marginal benefit upward (downward) both the optimal $x_{A,s}$ and A’s profits go up (down). ■

We are still in the proof of Proposition A.2. Part (i) and part (iii) follow immediately from combining the Lemma 1 and Lemma 2. For the first half of part (ii) notice that whatever the naive A actually does, increases in the actual costs make it according to Lemma 1 weakly less likely that his information will actually be sold to C. For the second half part (ii) notice that for $\hat{c}_{B,s} \leq H(0,1)$ the information is always sold and that according to Lemma 1 higher imagined costs make the naive A less cautious such that his true profits fall. For $\hat{c}_{B,s} > H(0,1)$ naive A’s behavior is constant in $\hat{c}_{B,n}$ and changes in imagined costs have no further effect on his actual payoff. For the first half of part (iv) notice first that B earns zero from the information sale in equilibrium regions IV and V, that is, when $\hat{c}_{B,s} \geq H(0,1)$. For lower $\hat{c}_{B,s}$ notice that his actual costs do not affect the amount of information that is provided. However, he obviously has to pay $c_{B,s}$. For the second half of part (iv) use Lemma 1 and (6) to see that A provides more information in response to an increase in $\hat{c}_{B,n}$, and notice that B does not actually pay the imagined cost $\hat{c}_{B,n}$, while his revenue increases. Finally, for part (v) observe that in equilibrium regions I, II, and III changes in $\hat{c}_{B,s}$ neither affect the
amount of information that players A will provide nor the probability that the true (low-cost) B offers $m_A$ for sale. In region V, B never sells, hence $\pi_C = 0$. In region IV, increasing costs imply that B sells with a decreasing probability. This increases the amount of information that the sophisticated A provides and hence C’s expected surplus conditional on buying. However, at the same time it becomes less likely that that surplus will materialize. Finally, for the last statement, $\frac{d\pi_C}{d\hat{c}_{B,n}} \geq 0$, observe that for $\hat{c}_{B,s} < H(0,1)$ higher imagined costs provide C with more information from the naive A while the information provision from the sophisticated A as well as B’s probability of selling remain constant, and that if $\hat{c}_{B,s} \geq H(0,1)$, both are constant in $\hat{c}_{B,n}$. ■