

« AI devices and liability »

Auteurs


Kene Boun My, Julien Jacob, Mathieu Lefebvre

Document de Travail n° 2024 – 24

Juin 2024

**Bureau d'Économie
Théorique et Appliquée
BETA**

www.beta-economics.fr

 @beta_economics

Contact :
jaoulgrammare@beta-cnrs.unistra.fr

AI devices and liability*

Kene Boun My[†], Julien Jacob[‡], Mathieu Lefebvre[§]

June 6, 2024

Abstract

We propose a new theoretical framework to analyze the incentives provided by different allocations of liability in the case of (semi)autonomous devices which are a source of risk of accident. We consider three key agents, an AI provider (scientist), a producer and a consumer, and look at the effect of different rules of sharing liability on the decision making of each type of agent. In addition we test the theoretical predictions in an original lab experiment. We show that liability on the scientist and the producer is efficient in reducing their misbehaviors. We also find that liability on the consumer increases her incentives to control the risk of an accident (in case of a semi-autonomous device). However, the absence of consumer's control (full autonomous device) and liability decreases the consumer's propensity to buy the good. We complete our study by making a social welfare analysis. It highlights the importance of letting the producer liable in order to provide the consumer with confidence in the technology, especially in the case of a full autonomy of the good.

Keywords: AI, Liability Sharing Rules, asymmetric information, lab experiment.

JEL Classification: C91, D82, K13, K32

*The authors thank participants at the workshop "Economie, Santé et Environnement : Risques et Incertitudes" at the University Paris-Saclay. This research has been conducted with the financial support of the IDEX Attractivité. This work was also supported by the French National Research Agency Grant ANR-17-EURE-0020, and by the Excellence Initiative of Aix-Marseille University - A*MIDEX. The authors are grateful to Maria-Candida Miguel and Julie Rabenandrasana for research assistance.

[†]University of Strasbourg, BETA, CNRS, Strasbourg, France, bounmy@unistra.fr.

[‡]University of Strasbourg, BETA, CNRS, Strasbourg, France, julienjacob@unistra.fr.

[§]University of Strasbourg, BETA, CNRS, Strasbourg, France, mathieu.lefebvre@unistra.fr.

1 Introduction

Artificial intelligence (AI) is rapidly emerging across diverse sectors, promising to deliver new value and reshape the employment landscape. A good example is the automotive industry (i.e. self-driving cars) that is undergoing significant transformation due to this technological evolution. Anticipated changes include a reduction in the risks of accidents even if complete elimination remains elusive. Moreover, the established system of liability for car accidents needs to be reevaluated in light of these advancements. Statistical evidence indicates that a substantial majority of car accidents (95%) can be attributed to human behavior.¹ While AI has the potential to mitigate the impact of human errors, it cannot entirely eradicate the risk of accidents. This is due to the persisting and inherent risk associated with machine malfunctions or design defects, as discussed in De Chiara et al. (2021) and Guerra et al. (2022a).²

The case of self-driving cars is instructive as in order to manage the evolving landscape of car accidents influenced by AI, public policy will play a crucial role.³ So far this policy approach incorporates two key instruments: *ex ante* market authorization and *ex post* liability. Market authorization ensures that products and technologies meet predefined quality standards before entering the market. Conversely, liability operates post-accident, compelling individuals to rectify any harm caused. Civil liability thus serves as a policy tool, incentivizing efforts to mitigate the risk of harm, as extensively explored in the literature on law and economics since seminal works by Calabresi (1970), Brown (1973), and Shavell (1980).

Despite the presence of some challenges in its implementation, liability is a particular interesting tool in the context of AI's impact since it allows to account for the dynamic nature of technological advancements. Unlike *ex ante* market authorization, which sets predefined standards before market entry, liability operates reactively, addressing incidents post-occurrence. This ex-post approach recognizes the evolving nature of technology and the potential for unforeseen challenges despite rigorous pre-market assessments. By prioritizing liability, the legal and regulatory framework could encourage a responsive and adaptive strategy. It would enable swift adjustments and corrective measures in the aftermath of accidents, fostering a more agile system capable of addressing emerging risks and complexities associated with AI. Additionally, one advantage of liability is that it is by nature in line with the fundamental principles of risk prevention and individual accountability, thus providing a robust incentive structure for entities to continually enhance their safety measures and control the risk of causing harms.

While the integration of new technologies in cars aims at liberating individuals from the task of controlling the vehicle, it introduces new challenges to risk mitigation policies, particularly concerning liability for car

¹For the US case, see the survey made by National Highway Transportation Safety Administration (NHTSA) of the U.S. Department of Transportation : <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812456>

²See also Commission (2020).

³We use the case of self-driving cars as a running example in this paper but most of the issues can be associated to other devices with embedded AI.

accidents. Presently, the foundation of liability for car accidents is rooted in the Vienna Convention (1968), designating the car driver as the *primary* responsible party due to their control over the vehicle. However, as autonomous vehicles aspire to emancipate the driver from active control, transforming them into passive users, this conventional paradigm becomes obsolete, which necessitates a comprehensive reassessment of the entire liability framework. Notably, no country has yet established a definitive approach to defining liability for accidents involving autonomous vehicles, as highlighted in Guerra et al. (2022b). Given the multifaceted influence of various actors on risk levels, many legal scholars advocate for a shared liability model among these entities.

Three primary actors come into focus: the AI provider, the car producer, and the user/driver. Different ways of sharing liability are currently under discussion in the context of autonomous vehicles. First, a model that allocates complete and sole liability to the car producer has been proposed, assuming the producer's expertise in controlling the reliability of the AI system through rigorous testing. This approach enables the car producer to ensure the safety and reliability of the vehicle without burdening the AI-provider with liability concerns, thereby fostering the emergence of these technologies (Elish and Hwang, 2015; Ilkova and A. Ilka, 2017; Kalra et al., 2009). An alternative model suggests sharing liability between the producer and the AI provider. This arrangement aims to provide additional incentives for the AI provider, acknowledging the potential informational asymmetry between the provider and the car producer (Vladeck, 2014; Gless et al., 2016; Tai, 2018). Finally a last model advocates for shared liability among the producer, the AI provider, and the user. This approach applies to non-completely autonomous vehicles, motivating the driver to exercise caution. In the case of fully autonomous vehicles, it serves to regulate vehicle use⁴ (Duffy and Hopkins, 2013; Schaefer et al., 2009; Kelley et al., 2010; Scherer, 2018; Tai, 2018; Gless et al., 2016).

The aim of this paper is to present new economic evidence within the ongoing discussion on the allocation of liability in the case of (semi)autonomous devices, with a focus on car risk accident management. Our analytical framework involves three key agents: an AI provider, a (car) producer (with embedded AI), and a consumer (user/driver). The AI provider engages in research and development (R&D) to design a technology that may carry defects. By intensifying R&D efforts, the AI provider enhances the likelihood of developing a more reliable technology. The AI provider, having perfect information about the technology's reliability, shares truly or falsely this information with the producer. The producer, unable to discern the technology type without incurring testing costs, incorporates it into a car offered to the consumer. The consumer, uninformed about the technology, decides whether or not to purchase the car on the basis on information provided by the producer. Exogenous selling prices vary, with car incorporating reliable technology commanding higher prices. We differentiate between semi-autonomous vehicles, requiring consumer efforts to reduce accident probabilities, and fully autonomous vehicles, where consumers lack control over the accident risk levels. The

⁴In Germany, it refers to the doctrine of *Betriebsgefahr*, see Janal (2016).

goal is to look at the effect of different liability rules on the decision making of each type of agent. In particular, we consider three rules: a full liability on car producer, a 50-50 sharing between the car producer and the AI provider, and an equal sharing between the AI provider, the car producer and the user/driver. In addition to the theoretical analysis we test the predictions in an original lab experiment. We adopt a similar environment in which subjects play either one of three different agents. We conduct different treatments according to the degree of liability and the autonomy of the vehicle. Our experimental design is well-suited to study how the apportion of liability among those different agents affect their behavior. In addition to measuring treatment effects, our approach also enables us to highlight eventual psychological factors that could hamper both the incentives to manage the risk (provided by the liability rule) and the adoption of the technology in the good/vehicle. Beyond the three liability rules, we test the consumers' behavior towards self-driving technology (and especially the decision to buy or not the car) by making a comparison between cases of autonomous and non-autonomous vehicles. Thus it explicitly takes into account different information asymmetries between the three agents that could affect efficient risk management and technological diffusion of self-driving technologies.

The theoretical framework shows that, in the presence of those three agents, liability on the scientist and the producer is efficient in reducing their misbehaviors, and liability on the consumer increases its incentives to control the risk (in case of a semi-autonomous device). The absence of consumer's control and allocating liability to the consumer decrease the consumer's propensity to buy the car. Results from the experiment show that the degree of sharing liability has indeed an effect on each agent's decision and that this effect depends on the autonomy (or its absence) of the technology. In particular, liability increases the consumer's effort to avoid a damage but decreases her propensity to buy the good/car. Liability also decreases the proportion of lies by the scientist and the producer. The autonomy of the technology does not appear to play a role except for the producer who is more willing to know the true quality of the technology when it is autonomous. Overall, the results show that liability is important to avoid dishonest behavior and push users to make effort to avoid damages.

The rest of the paper is organized as follows. In Section 2, we highlight our contributions to the literature. In Section 3, we present our theoretical framework. Section 4 presents the experimental design and the predictions that we test with this experiment. In Section 5, we present the results and in Section 6, we discuss them and conclude.

2 Contributions to the literature

Our paper contributes to the literature on several grounds. The sharing of liability in the case of accidents involving autonomous vehicles is currently under debate among lawyers and lawmakers. In particular, the

legal scholars have extensively discussed the three rules of apportionment of liability that we discuss in our paper. The law and economics literature has also proposed theoretical as well as empirical analysis of the economic consequences of civil liability. More recently, the literature in behavioral economics has looked at the reporting of private signals in asymmetric information settings and in particular at dishonest behavior.

Legal literature

The legal literature on liability for (semi)autonomous cars is very large and discuss many challenges in both implementing and sharing liability rules.⁵ The implementation of liability rules in case of accidents involving vehicles assisted by AI devices could be a challenge. As noted by Scherer (2016), the autonomy of AI devices could break the causal relationship between the AI programming (made by the AI provider) and the malfunctioning of the AI that leads to the accident; thus shielding the AI provider from liability. To circumvent this problem, some legal scholars advocate for a liability on the AI devices itself, by giving it a new kind of personality (Solum, 1992; Hilgendorf, 2014; Vladeck, 2014; Rothenberg, 2016). But this option is highly criticized (Lopucki, 2017; Rothenberg, 2016; Nevejans, 2016; Brown, 2021), and some scholars note that liability could still be devoted to the AI provider in the frame of a global risk management strategy (Gless et al., 2016; Chesterman, 2020)⁶. As noted above, the possibility to put liability to the vehicle manufacturer instead of the AI provider is also discussed. This option is supported by a need to decrease the AI provider's risk to encourage innovation and development of new technologies but also to ensure sufficient quality and technical control by the producers (Elish and Hwang, 2015; Ilkova and A. Ilka, 2017; Kalra et al., 2009). Finally, (partial) liability for the user has also been proposed (Duffy and Hopkins, 2013; Schaerer et al., 2009; Kelley et al., 2010; Scherer, 2018; Tai, 2018; Gless et al., 2016) and this even in the case of fully autonomous vehicles.⁷ Liability to the (passive) user could be enforced because of the need to regulate the intensity of the use of the car, similarly to strict liability applying to the owners of animals in most countries.

Law and economics literature

Although legal scholars may rely on economic arguments to help supporting their doctrinal positions (like the need to foster innovation, or the need to provide incentives to control the risk), they do not introduce (normative) economic assessment of the different possibilities of sharing liability. Emerging research in law and economics have begun to study this question.

One important aspect of the problem is to define who has control on the risk of an accident. Shavell (2020) considers the case of accidents involving two autonomous vehicles (where both agents are injurer and victim simultaneously) but does not distinguish between the AI provider and the car manufacturer. He also considers the possibility that both the driver and the car producer have control on the level of risk. The driver can control

⁵See Gless et al. (2016); Vladeck (2014); Tai (2018) or Briquet et al. (2024) for surveys on this topic.

⁶This includes, for example, the obligation to continuously monitor customer feedback, to react immediately to complaints of defects or accidents (by updating the software, etc.).

⁷In case of semi-autonomous cars, for which the user is still a driver having control on the vehicle, liability for the driver follows the same rationale as those of the Vienna convention.

the level of activity, through the miles traveled. The car manufacturer can also affect the risk through a better care to the production of the vehicle. Shavell (2020) shows that optimal incentives are derived from strict liability on the producer for the combined harm, augmented by a mileage fee on users. Guerra et al. (2022a) go beyond the case of autonomous vehicles and consider a three-agents framework with a manufacturer, an operator, and a victim. The three agents have an impact on the level of risk, via R&D (manufacturer) or care and activity levels (operator and victims), which are all substitutables. The manufacturer is both the producer and the seller of the good which is used by the operator. Guerra et al. (2022a) show that a second-best liability scheme is to make operator and victim liable for negligence, and the manufacturer residually liable only in case of non-negligence from operator and victim. In that case R&D and care levels are optimal but activity levels are not controlled. However, these studies do not take into account the transaction between the user/consumer and the producer. In our framework, we introduce the decision making of the user that can decide to buy or not an AI embedded good depending on different characteristics of the good (quality and/or autonomy). This affects also the possible sharing of liability. In addition, we distinguish the AI provider from the manufacturer (and seller) of the good, both having an impact on the level of the risk through actions which are not substitutables.

Another important issue is related to the full or partial autonomy of the vehicle. Talley (2019) specifically addresses the coexistence of autonomous and non-autonomous vehicles when the user has the choice to rely on one or the other before entering traffic. In case of non-autonomous vehicles, the user has to make an effort to reduce the risk. In case of autonomous car, the user *ex ante* chooses a level of R&D which determines the extent of pre-programmed scenarios (of accidents) which can be successfully managed by the AI. If the autonomous car meets a scenario of accident which is not pre-programmed in its database, an accident occurs. Knowing that autonomous cars perfectly manage the pre-programmed scenarios, (potential) victims may “outsmart” the self-driving technology, thereby taking fewer precautions themselves. Talley (2019) shows that optimal incentives are provided by a strict liability on the car user with a defense for contributory negligence of the victim. An originality of this paper is to begin a reflection on the coexistence of both car technologies (autonomous, and non autonomous), here in the case of interactions between cars and other agents (e.g., pedestrians). But this first analysis is made under simplifying assumptions, especially the presence of only one agent who makes all decisions relative to the driving of the car (ex ante R&D in case of self-driving car or care level in case of non-autonomous vehicle).

Close to our framework, De Chiara et al. (2021) consider a framework wherein consumers can buy either an autonomous car provided by a monopolistic manufacturer, or a non-autonomous car available on a perfectly competitive market. Consumers differ in their cost of attention (care) to control the risk of accident when using a non-autonomous car. The benefit of buying an autonomous car lies in the fact of not making any effort in care and not being liable in case of an accident. The manufacturer of the autonomous car has to

decide about the selling price, an amount in R&D (fixed cost) that will reduce the cost of post-marketing care, and a level of a post-marketing care (fixed cost) that allows to reduce the probability of each sold car to cause an accident. Thus De Chiara et al. (2021) assume that the single manufacturer of autonomous cars has the control on their dangerousness and is totally liable in case of accident. Comparing strict liability and negligence, they show that both liability rules can lead to optimal levels of care, but only strict liability on the manufacturer of autonomous car leads to efficient R&D investments and favors the adoption of autonomous cars. In a second version of the model, they consider the case where the probability of accident also depends on the level of activity (miles traveled) and show that strict liability for the manufacturer is optimal only if the users of autonomous cars can be fined in case of an accident. Contrary to our study, De Chiara et al. (2021) do not compare different ways of sharing liability in case of accidents involving autonomous cars: the unique manufacturer of autonomous car is always fully liable in case of accident. Also both technologies coexist and the efficiency of all investments in R&D and efforts in care is certain. Furthermore all information about the quality of cars are common knowledge. Letting the consumer to buy or not a car is a way to measure the diffusion of self-driving technologies.

All the previous papers propose a comparison between a strict liability rule and a negligence rule devoted to one or the other agent, possibly completed by another policy tool (like a payment to a third-party for example). Our paper complements these studies by looking at the incentives provided by three different sharing of liability that are currently discussed by legal scholars and law makers. We focus on accidents towards external victims and contrary to the other contributions, our analysis is both theoretical and experimental which allows to test empirically the model predictions.

Behavioral economics literature

Our paper is also somehow related to a growing literature in behavioral economics (but also in psychology and sociology) that deals with trust and dishonest behavior. Indeed in our framework of analysis, the AI provider and the producer both have the opportunity to lie about the nature of the technology embedded in the good (car). In classic economic theory, it is often assumed that people are willing to misreport private information if the material incentives of acting dishonestly outweigh those of acting honestly. Saying differently, the individual engage in dishonesty whenever this behavior pays off. However, although reporting private information in asymmetric information situations is common to many economic activities, empirical works on the field and in the lab have shown that people often behave otherwise. Even when there is no scrutiny, no negative externalities to lying and lies are profitable, there are individuals that do not lie (Tergiman and Villeval, 2022). Thus several theoretical contributions have departed from the individual payoffs maximizing assumption to take into account the preference for truth-telling (Kartik, 2009; Matsushima, 2008; Kartik et al., 2014).

In a recent meta-analysis on 72 experimental studies, Abeler et al. (2019) have shown that subjects forgo

on average about three-quarters of the potential gains from lying, which is a strong departure from the standard economic prediction. This preference for truth-telling is robust to changing the payoff level or repeating the decisions. There are several reasons that can explain these results. It has been shown that individuals may suffer from lying aversion (see, e.g., Ellingsen and Johannesson, 2004; Hurkens and Kartik, 2009; Fischbacher and Föllmi-Heusi, 2013) or care about the reputational cost of lying (Kajackaite and Gneezy, 2017; Dufwenberg and Dufwenberg, 2018; Khalmetski and Sliwka, 2019). The role of social norms and social comparison has also been pointed out, in particular guilt aversion (Charness and Dufwenberg, 2006).

Although our paper is not about the determinants of lying, our framework contributes to identify situations of dishonesty and in particular how liability rules may discipline behavior. As pointed out by Tergiman and Villeval (2022): "the literature on credence goods has shown that among reputation, verifiability, liability, competition and the interaction thereamong, only liability (the obligation for the seller to provide sufficient quality) leads to significantly more honesty" (see i.e. Balafoutas et al., 2013; Beck et al., 2014; Mimra et al., 2016; Feltovich, 2019). In our case we provide new insight on the effects of sharing liability when multiple agents are concerned.

3 Theory

3.1 Basics

We consider an economy comprising 3 risk-neutral agents: a Scientist (S), a Producer (P), and a Consumer (C). They are all independent decision-makers, and they are all endowed with a wealth W_i ($i = S, P, C$). The Consumer can buy a good (e.g. a car) from the Producer. In order to make the car, the Producer needs a technology provided by the Scientist. In our framework, the Scientist is therefore the AI provider, to use the same terminology as before. In the case where the Consumer buys the car, an accident can occur during its use. In such a case, a harm H is caused to a passive third-party. The risk of causing harm depends on several factors.

First, the Scientist can design either a technology of good quality (G-quality hereafter, indexed by G), which leads to a probability p_G of causing harm, or a technology of bad quality (B-quality hereafter, indexed by B), which leads to a probability p_B of causing harm; with $0 < p_G < p_B < 1$. In order to decrease the likelihood of designing a technology of B-quality, the Scientist can make an effort e , for a cost $c(e)$, with $e \geq 0$, $c'(e) > 0$, $c''(e) > 0$, $c(0) = 0$. A higher effort e decreases the probability $p(e)$ of designing a technology of B-quality, with $0 < p(e) < 1$, $p'(e) < 0$, $p''(e) \geq 0$.

Second, the technology embedded in the car can be "Autonomous" or "Non-Autonomous". When using it,

the Consumer has no control on a car embedded with an Autonomous technology. But in the case on a Non-Autonomous technology, the Consumer has a control on the probability of causing harm. The Consumer can make an effort, denoted by ϵ ($\epsilon \geq 0$), to decrease the probability that a harm occurs, that is to reduce p_B or p_G (depending on the quality of the technology). We assume that a level of effort ϵ gives a disutility $d(\epsilon)$ which does not reduce the consumer's level of wealth W_C ; with $d(0) = 0$, $d'(\epsilon) > 0$, $d''(\epsilon) > 0$. As a consequence, in the presence of a Non-Autonomous technology, the probability that a technology of B-quality causes harm is $p_B(\epsilon)$, with $p_B(\epsilon = 0) = p_B$ and $p_B(\epsilon > 0) < p_B$ (with $p'_B(\epsilon) < 0$, $p''_B(\epsilon) \geq 0$). Similarly, for a technology of G-quality, we have a probability $p_G(\epsilon)$ of causing harm, with $p_G(\epsilon = 0) = p_G$ and $p_G(\epsilon > 0) < p_G$ (with $p'_G(\epsilon) < 0$, $p''_G(\epsilon) \geq 0$).⁸

3.2 Timing

The embedded technology in the car is either Autonomous or a Non-Autonomous and the three agents make decisions sequentially. First, the Scientist decides the amount of effort e to decrease the probability of designing a B-quality technology. Nature then chooses the quality of the technology (B or G), according to the probability $p(e)$. The Scientist observes whether the technology is of G-quality, or of B-quality. The Scientist transfers the technology to the Producer.⁹ The Scientist declares the quality of the technology to the Producer. The quality is unobservable to the Producer unless she invests an amount $K > 0$ in information acquisition. This means that the Scientist can lie to the Producer about the quality of the technology.

Second, the Producer offers the the car to the Consumer with the technology embedded. The Consumer knows that two qualities of technologies exist and they can cause harm to a passive third-party with different probabilities (p_G or p_B in the Autonomous case and $p_G(\epsilon)$ and $p_B(\epsilon)$ in the Non-Autonomous case) but is unable to observe the quality of the embedded technology. Indeed the Producer announces a quality of the technology to the Consumer and the price varies accordingly. The price is ρ_G for a G-quality technology and ρ_B for a B-quality one; with $\rho_G > \rho_B > 0$. If the consumer buys the good, the gains from sale are shared between the Producer and the Scientist with a share α for the Producer and a share $(1 - \alpha)$ for the Scientist. Finally, given the information provided by the Producer, the Consumer decides whether to buy the car, or not. In the case where the Consumer buys the car, Nature then decides whether a harm H is caused, or not. Recall that in the case of a car embedded with an Autonomous technology, the Consumer has no control on the probability of causing harm (which is common knowledge). By contrast, in the case of a Non-Autonomous

⁸As an illustration, we can consider the example of fully or partially autonomous vehicles. In the case of a full autonomous car, if a deficiency occurs during the journey, the car driver has no control, and cannot try to avoid the accident. In that case, p_j is both the probability of deficiency and of causing an accident. In the case of a non-autonomous car, if a deficiency occurs (with probability p_j) the driver can take back control of the car to try to avoid the accident. It all depends on how vigilant she is (expressed by the effort ϵ_j , which reduces the level of $p_j(\cdot)$). The vigilance of the driver makes possible to decrease further the probability of having an accident.

⁹To fit in with the experiment we are conducting below on the basis of that model, we suppose no monetary transaction between the Scientist and the Producer. Such an assumption does not reduce the extent of our results.

technology, the Consumer decides about an effort ϵ to decrease further the probability of causing harm. If no accident occurs, the Consumer earns a benefit B_{sup} from using the car. If an accident occurs, the consumer earns B_{inf} , with $B_{sup} > B_{inf} > 0$. Equilibrium decisions are determined by backward induction.

3.3 First-best

Before determining the decentralized equilibria, we first derive the first-best decisions. A normative benchmark is obtained by determining the decisions that would be made by a benevolent, omniscient and omnipotent dictator who aims at maximizing the sum of all benefits and costs earned/incurred by the whole Society from designing and consuming the good under consideration (the car).

The benevolent dictator aims at maximizing:

$$SW(e) = \sum_i W_i + B_{sup} - c(e) - [p(e)p_B + (1 - p(e))p_G](H + (B_{sup} - B_{inf})) \quad (1)$$

with $\sum_i W_i = W_S + W_P + W_C$, in the case where an Autonomous technology can be designed, and

$$SW(e, \epsilon_B, \epsilon_G) = \sum_i W_i + B_{sup} - c(e) - p(e)d(\epsilon_B) - (1 - p(e))d(\epsilon_G) - [p(e)p_B(\epsilon_B) + (1 - p(e))p_G(\epsilon_G)](H + (B_{sup} - B_{inf})) \quad (2)$$

in the case where a Non-Autonomous technology can be designed. We pose: $B_{sup} - c(0) - d(\epsilon_B = 0) - p_B(\epsilon_B = 0)(H + (B_{sup} - B_{inf})) > 0$, which means that in the case of a car embedded with an Non-Autonomous technology of B-quality, the use of the car is socially desirable. The same applies with an Autonomous technology since $p_B(\epsilon_B = 0) = p_B$. ϵ_B and ϵ_G are efforts aiming at reducing $p_B(\epsilon)$ and $p_G(\epsilon)$ respectively. These efforts are conditional to the quality of technology but the dictator has full information and observes the quality of the technology.

In the case of an Autonomous technology, first-best effort in e , designated e_A^{**} hereafter, satisfies:

$$\frac{\partial SW(e)}{\partial e} = 0 \Rightarrow -p'(e_A^{**})(p_B - p_G)(H + (B_{sup} - B_{inf})) = c'(e_A^{**}) \quad (3)$$

with the subscript A meaning ‘‘Autonomous’’.

In the case of a Non-Autonomous technology, first-best efforts in e , ϵ_B and ϵ_G , designated e_{NA}^{**} and ϵ_B^{**} , ϵ_G^{**} respectively hereafter, satisfy:

$$\frac{\partial SW(\epsilon_B)}{\partial \epsilon_B} = 0 \Rightarrow -p'_B(\epsilon_B^{**})(H + (B_{sup} - B_{inf})) = d'(\epsilon_B^{**}) \quad (4)$$

and

$$\frac{\partial SW(\epsilon_G)}{\partial \epsilon_G} = 0 \Rightarrow -p'_G(\epsilon_G^{**})(H + (B_{sup} - B_{inf})) = d'(\epsilon_G^{**})(5)$$

and finally

$$\frac{\partial SW(e, \epsilon_B, \epsilon_G)}{\partial e} = 0 \Rightarrow -p'(e_{NA}^{**})(p_B(\epsilon_B^{**}) - p_G(\epsilon_G^{**}))(H + (B_{sup} - B_{inf})) = c'(e_{NA}^{**})(6)$$

with the subscript NA meaning “Non-Autonomous”.

Following the sequence of individual decisions (as described above), the decision about ϵ is conditional to the observation of the quality of the technology. Solving backward, the decision about e is made by taking into account the (subsequent) conditional decisions about ϵ_j^{**} , $j = B, G$.

In this first-best case, the dictator has complete information on the quality of the technology, and decisions are always made in accordance with the true state of Nature (there is no lie). So there is no need to seek for information about the quality of the technology. Also, there is no selling prices, since the transaction between the Consumer and the Producer is only an internal transfer (at the scale of the whole Society). We finally suppose that a G-quality technology provides a higher social welfare than a B-quality, for both the Autonomous and the Non-Autonomous cases.¹⁰

3.4 Private equilibria

When decisions are made by private agents (Scientist, Producer, and Consumer), civil liability applies in case of an accident. We suppose that liability is strict (see Shavell, 1980), and damages are compensatory. As a consequence, whenever a harm occurs, the amount H has to be repaired (compensation of the passive third-party victim). The payment of H is shared between private agents: each agent has to pay a share l_i of H , with $i = S, P, C$ and $0 \leq l_i \leq 1$, $l_S + l_P + l_C = 1$. So, in absolute value, each agent i has to pay an amount $l_i H$ in case of accident.¹¹

Applying backward induction, first the Consumer has to decide to buy (or not) the car. But in the case of a car embedded with a Non-Autonomous technology, the consumer has to decide about the level of effort ϵ to decrease the probability of causing harm. In case of an embedded Autonomous technology, the Consumer decides to buy the car offered by the Producer (with quality $j = B, G$) if the following condition is satisfied:

$$U_C = B_{sup} - \rho_j - [\phi_C(G|j)p_G + (1 - \phi_C(G|j))p_B](l_C H + (B_{sup} - B_{inf})) > 0 \quad (7)$$

¹⁰This requires: $(p_B(\epsilon_B^{**}) - p_G(\epsilon_G^{**}))(H + (B_{sup} - B_{inf})) - (d(\epsilon_G^{**}) - d(\epsilon_B^{**})) - c(e_{NA}^{**}) > 0$ for the Non-Autonomous case. We assume this condition to be satisfied.

¹¹In practice, liability is limited by solvency constraint. For an agent i , if NW_i is her net wealth (i.e., its wealth W_i , net of any other monetary expense), then the amount in damages to pay in case of harm should be $\min\{NW_i, l_i H\}$. In our analysis, we make the simplifying assumption that there is no insolvency issue. This means that all agents are able to pay for their share of liability: $NW_i > l_i H$.

with $\phi_C(G|j) \in [0,1]$ the subjective probability, for the Consumer, that the technology embedded is of G-quality knowing that the Producer has declared the quality to be j , with $j = B, G$.¹²

In case of a car embedded with a Non-Autonomous technology, the Consumer decides to buy the car offered by the Producer (quality $j = B, G$ declared) if the following condition is satisfied:

$$U_C(\epsilon) = B_{sup} - \rho_j - d(\epsilon^*) - [\phi_C(G|j)p_G(\epsilon^*) + (1 - \phi_C(G|j))p_B(\epsilon^*)] (l_C H + (B_{sup} - B_{inf})) > 0 \quad (8)$$

with ϵ^* the equilibrium level of effort, in reducing the probabilities $p_j(\epsilon)$ of causing harm.

The effort ϵ^* satisfies:

$$\frac{\partial U_C(\epsilon)}{\partial \epsilon} = 0 \Rightarrow -[\phi_C(G|j)p'_G(\epsilon^*) + (1 - \phi_C(G|j))p'_B(\epsilon^*)] (l_C H + (B_{sup} - B_{inf})) = d'(\epsilon^*) \quad (9)$$

Comparing (4) and (5) with (9) makes us possible to deduce the following Proposition.

Proposition 1 *Consider the Consumer buys a car embedded with a Non-Autonomous technology, and has an imperfect belief on the quality of the technology embedded in the car (B or G).*

(i) *In the case where the Consumer's effort ϵ has the same efficacy whatever the quality of the technology (i.e., $p'_B(\epsilon) = p'_G(\epsilon)$, ϵ given), then her effort is socially optimal if she faces full liability in case of accident (i.e., $\epsilon^* = \epsilon_G^{**} = \epsilon_B^{**}$ when $l_C = 1$). But if the Consumer faces partial liability ($l_C < 1$), then the level of effort is lower than the optimal one.*

(ii) *In the case where the efficacy of the Consumer's effort differs depending on the quality of the technology (i.e. $-p'_j(\epsilon) \neq -p'_{-j}(\epsilon)$), full liability of the Consumer does not lead her to make an optimal effort.*

The Consumer always internalizes the disutility of making effort ϵ , but her private (marginal) utility from effort does not always fit with the social one. This is the case when the Consumer's effort efficacy is the same whatever the technology used, and when full liability apply. However, if the Consumer's effort has not the same efficacy whatever the quality of the technology, to the extent that the Consumer has an imperfect belief on the *true* quality, her effort will diverge from the optimal level despite a full liability. This also applies with partial liability.¹³

Proposition 2 *Increasing the burden of liability on the Consumer (i.e., increasing l_C) decreases the interest for the Consumer to buy the car. But conditional to the purchase of the car, increasing the level of liability of the Consumer increases the level of effort ϵ .*

¹²In other words, these subjective probabilities represent the trust of the Consumer in the declaration of the Producer. $(1 - \phi_C(G|j))$ is thus the subjective probability of the technology to be B-quality knowing j is declared.

¹³Note that even in a case of partial liability, the effort made by the Consumer can be *higher* than the optimal one if she believes that the quality of the technology announced by the provider is the true one that provides the highest efficacy in effort. The Consumer thus provides a high level of effort, while the true benefit from effort is, in reality, low because of the low efficacy of effort. Incentives to make effort may *wrongly* be too strong if the Consumer is mistaking about the technology, even in the presence of partial liability.

Let us now turn to the decisions made by the Producer. The Producer first decides whether to invest in information seeking or not (in order to know the true quality of the technology), and then decides which quality j to announce to the Consumer ($j = B, G$). In case of accident, the Producer has a liability of $l_P H$. Moreover, following the accident an investigation is made by the Judge, which permits to discover the true quality of the technology embedded in the car. If, after an accident occurring, the investigation reveals that the Producer has not declared the technology to be of B-quality while it is the *true* quality, the producer can be charged with a fine $F_P \geq 0$. In the experiment below, we test the effect of the presence ($F_P > 0$) or the absence ($F_P = 0$) of a fine.

Let us first consider the decision about which quality to declare to the Consumer ($j = B, G$). Consider that the Producer believes that j is the true quality of the technology. He declares a quality j in accordance with his belief j (and thus she does not declare the opposite quality, $-j$) if

$$\alpha \rho_j - p_j(\epsilon_j^*) l_P H > \alpha \rho_{-j} - p_j(\epsilon_{-j}^*) l_P H \quad (10)$$

in the case of a Non-Autonomous technology,¹⁴ and:

$$\alpha \rho_j - p_j l_P H > \alpha \rho_{-j} - p_j l_P H \quad (11)$$

in the case of an Autonomous technology. For the special case where there is possibility to be fined when the B-quality is not released to the Consumer (in case of an Autonomous technology), the Producer prefers to declare the technology to be B-quality (if he thinks that B is the true quality) when: $\alpha \rho_B - p_B(l_P H) > \alpha \rho_G - p_B(l_P H + F_P)$.¹⁵

Looking at the decision to pay a cost $K > 0$ in order to know with certainty the true quality of the technology, when the Scientist declares quality j , the Producer pays K if:

$$\phi_P(-j|j) [\alpha(\rho_{-j} - \rho_j) - (p_{-j}(\epsilon_{-j}^*) - p_j(\epsilon_j^*)) l_P H] > K \quad (12)$$

in the case of a Non-Autonomous technology, and:

$$\begin{aligned} \phi_P(-j|j) [\alpha(\rho_{-j} - \rho_j) - (p_{-j} - p_j) l_P H] &> K \\ \Rightarrow \phi_P(-j|j) \alpha(\rho_{-j} - \rho_j) &> K \end{aligned} \quad (13)$$

¹⁴When declaring the quality j , we suppose the Producer expects the Consumer to make an effort in accordance with this quality ϵ_j^* .

¹⁵Of course this reasoning suppose that (7) and (8) are verified (respectively); the Consumer buys the car given her level of trust in the announcement made by the Producer.

in the case of a Autonomous technology. For the special case where there is possibility to be fined when the B-quality is not released to the Consumer (in case of an Autonomous car), the Producer invests K when $\phi_P(B|G) [\alpha(\rho_B - \rho_G) + p_B F_P] > K$. $\phi_P(-j|j)$ is the Producer's belief in the true quality to be $-j$ when the Scientist declares type j . We remark that in case of a Non-Autonomous technology (Eq. 12), both the differential in selling prices, $\alpha(\rho_{-j} - \rho_j)$, and the differential in expected liability, $(p_{-j}(\epsilon_{-j}^*) - p_{-j}(\epsilon_j^*))l_P H$, take part in the decision. In the case of an Autonomous technology (Eq. 13), only the differential in prices takes part. This is so because in the former case, the Consumer has an impact on the level of the risk. Thus the Producer's announcement alters the Consumer's decision, which in turn alters the level of risk for the Producer. By contrast, in the presence of a car endowed with an Autonomous technology, the risk of harm is exogenous. The Producer's announcement has no impact on the risk, and so it does not alter the decision to pay for information.

Proposition 3 *About the Producer's decisions*

(i) *In case of a Non-Autonomous technology, because $\rho_G > \rho_B$, if the Producer bears no liability (i.e., $l_P = 0$), the Producer lies to the Consumer when the technology is B-quality. However for a given ϵ and because of $p_G(\epsilon) < p_B(\epsilon)$, the higher the Producer's liability, the lower the incentives to lie and the higher the incentives to pay for information about the quality of the technology.*

(ii) *In case of an Autonomous technology, the Producer always lies to the Consumer when the technology is B-quality. As a consequence, The Producer has no incentives to pay for information about the quality of the technology. Only a sufficiently high fine (in case of accident) for not disclosing the true quality may deter the Producer to lie, and provide incentives to pay for information about the quality of the technology.*

(iii) *The Producer has no interest in lying when the technology is G-quality (whatever Autonomous or Non-Autonomous).*

(iv) *The Producer has no interest in paying for information when the Scientist announces a B-quality technology.*

Finally, the Scientist makes two decisions. First he decide about the effort e to spend to reduce the probability of designing a B-quality technology, and second he decides which quality of technology to announce to the Producer. In case of accident, the Scientist faces a liability of $l_S H$.

Let us first consider the decision about what type j (B or G) of quality to announce to the Producer. The Scientist observes the true quality j and decides about to announce it to the Producer (i.e., to declare j , and not the opposite quality $-j$) if:

$$(1 - \alpha)\rho_j - p_j(\epsilon_j^*)l_S H > (1 - \alpha)\rho_{-j} - p_j(\epsilon_{-j}^*)l_S H \quad (14)$$

in the case of a Non-Autonomous technology (and if the Producer announces to the Consumer the quality announced by the Scientist). In case of an Autonomous technology this condition is:

$$(1 - \alpha)\rho_j - p_j l_S H > (1 - \alpha)\rho_{-j} - p_j l_S H \quad (15)$$

Finally, we have to determine the amount e^* that the Scientist invests in order to decrease the likelihood to obtain a B-quality technology. This effort e^* satisfies:

$$\max_e E[\Pi_S] = p(e)\Pi_S(B - quality) + (1 - p(e))\Pi_S(G - quality) - c(e) \quad (16)$$

$E[\Pi_S]$ denoting the Scientist's expected profit, with:

$$\Pi_S(B - quality) = (1 - \alpha)\rho_B - p_B(\epsilon_B^*)l_S H$$

and

$$\Pi_S(G - quality) = (1 - \alpha)\rho_G - p_G(\epsilon_G^*)l_S H$$

in the case where there is no lie between the Scientist and the Producer, and no lie between the Producer and the Consumer.¹⁶

The resulting level of effort e^* satisfies:

$$\frac{\partial E[\Pi_S]}{\partial e} = 0 \Rightarrow -p'(e^*) [\Pi_S(G - quality) - \Pi_S(B - quality)] = c'(e^*) \quad (17)$$

Proposition 4 *About the Scientist's decisions*

(i) *In case of a Non-Autonomous technology, because $\rho_G > \rho_B$, if the Scientist bears no liability (i.e., $l_S = 0$), the Scientist lies to the Producer when he gets a B-quality technology. However, for a given ϵ , because $p_G(\epsilon) < p_B(\epsilon)$, the higher the Scientist's liability, the lower the incentives to lie.*

(ii) *In case of an Autonomous technology, the Scientist always lies to the Producer (whatever the degree of liability).*

(iii) *The Scientist has no interest in lying when he obtains a G-quality technology (whatever Autonomous or Non-Autonomous).*

(iv) *The level of effort e^* is always lower than the first-best level.*

Point (ii) of Propositions 3 and 4 rely on a similar rationale. In case of an Autonomous technology, once the technology is designed, the level of risk cannot be reduced by other decisions. The risk being exogenous,

¹⁶Here, both $\Pi_S(B - quality)$ and $\Pi_S(G - quality)$ are expressed for the case of Non-Autonomous technology. In case of an Autonomous technology, probabilities of harm are fixed to p_B and p_G for qualities B and G respectively.

profit-maximizing agents have thus incentives to always declare a G-quality technology to maximize earnings. Information is useless for the Producer, except in the case where he can be fined for miscommunication about the quality of the technology towards the Consumer.

Point (i) of Propositions 3 and 4 rely on a similar rationale. From Point (ii), we know that it is only the perspective of decreasing the level of risk (of paying damages) that provides both the Scientist and the Producer incentives to announce the true quality of the technology and for the Producer to pay to know the true quality. This incentive is increasing with liability. However, because of liability sharing among the agents ($\sum_l l_i = 1$), there is a trade-off in the incentives to provide.

According to Proposition 4 (iv), even with a full liability, the Scientist has no incentives in providing a level of effort e up to the socially optimal level since the Scientist does not capture the entire social benefit from this effort (and especially the Consumer's surplus from using the car). Proposition 3 (iv) is obvious since a B-quality technology provides the lowest sales revenue.

4 Experimental design

The experiment consists of a repeated game played by groups of three subjects for 5 rounds. The composition of each group is randomly changed every round and each participant may encounter any other participant only once. To avoid other confounders, the experiment is decontextualised but it is based on the theoretical framework exposed above. We use the term good instead of car. Each subject has a specific role, namely Scientist, Producer or Consumer. In each period, the three types of player make their decisions sequentially. The complete instructions are presented in Appendix A.3, and details about the calibrations are provided in Appendix A.5.

4.1 The sequence

Each round follows the same sequence. First the scientist makes decisions, then the producer and finally the consumer.

The scientist (S) receives an endowment of 1000 ECUS to develop a technology, which can be a G-quality or a B-quality according to a certain probability. In each round, the Scientist has to decide how much to invest in R&D to increase the probability of developing a G-quality technology. Table 1 shows the probability of obtaining a G-quality technology as a function of the amount invested by the Scientist. The bigger the investment in developing the technology, the more reliable it is likely to be.¹⁷

The difference between G and B technology lies in the probability of the technology to fail once it is used

¹⁷The highest investment exhausts completely the scientist's initial resources which makes possible a loss. All depends whether the consumer buy or not the good associated with this technology. See below.

Table 1: Probability of having a Good technology

Investment by the Scientist	0	4	16	64	128	512	1024
Probability of G-quality	10%	30%	40%	50%	60%	70%	80%

by the consumer, see below. After having made the decision, the scientist is informed of the nature of the technology and transfers the technology to the producer. To this end, the scientist must tell the producer whether the technology is G or B without obligation to tell the truth.

The producer (P) must then incorporate this technology into a good (i.e. a car in our theoretical framework) that is offered to the consumer. At the beginning of each period, the producer receives 1500 ECUS and can find out the true nature of the technology by paying a cost of 50 ECUs. In order to sell the good to the consumer, the producer has to announce the quality of the technology (either G or B). As for the scientist, the producer can announce any quality (and thus can lie). A good with a G-quality technology is sold 1500 ECUS while a good with a B-quality technology is sold at 1400 ECUs.

The consumer (C) receives 1500 ECUS and has to decide whether or not to buy the good offered by the producer. As explained above, the price of the good depends on the type of technology that is announced by the producer (and thus not its very nature) but its use benefits to the consumer. The benefit depends on the occurrence of a failure of the technology. The benefit for the consumer is 1800 ECUS, when no failure occurs while it is only 1200 ECU in case of failure. Indeed, in the absence of control by the consumer, a B-quality technology has 60% chance of failure, while a G-quality technology has 20% chance of failure.

4.2 Treatment conditions

We implement seven different treatment conditions in a between-subjects design. The treatment conditions differ on three dimensions. First, the technology can be autonomous, in a way that the consumer has no control on it, or non-autonomous and the consumer can control it. In the case of a non-autonomous technology (indexed NA), the consumer can make an effort that will reduce the likelihood of the technology failing. To do so, subjects worked on a tedious task : counting the number of 1s in a series of tables. Each table consisted of 50 randomly ordered 0s and 1s. This task did not require any prior knowledge and performance was easily measurable. Furthermore, there was little learning possibility and effort was costly in terms of disutility. In this experiment we do not consider a monetary cost of the effort but rather a cognitive cost. Which is probably more related to the reality if one consider the effort of vigilance in a self-driving car. Subjects have one minute to count correctly up to four tables. Table 2 shows how the effort made by the consumer can reduce the probability of a failure. In the case of an autonomous technology (indexed A), they cannot affect the probability of failure.

Table 2: Probability of failure according to consumer's effort

# of tables	0	1	2	3	4
G-quality technology	0.2	0.15	0.11	0.075	0.05
B-quality technology	0.6	0.45	0.33	0.23	0.15

Second, within each type of technology (autonomous or not), we differentiate between three ways of sharing liability. Indeed, in the case of technology's failure, it causes damage that must be borne by one or other agent, depending on the treatment. The total cost of the damage is 2400 ECUs and in treatments indexed P the producer is fully responsible to pay for the damage. In treatments indexed S, the cost of the damage is shared by the producer and the scientist and in treatments indexed C it is shared between the three agents. Finally we introduce a seventh treatment condition in which we introduce in the case of an autonomous technology, in addition to full liability to the producer, a financial fine of 500 ECUs in case of technology's failure and producer lies about the quality of the technology. The treatment is indexed F. Table 3 summarizes the different treatment conditions and their characteristics.

Table 3: Treatment conditions

Liability \ Technology	Autonomous	Non-autonomous
	P (Total)	T-AP
P (Total) + fine	T-AF	
P (1/2) and S (1/2)	T-AS	T-NAS
P (1/3), S (1/3) and C (1/3)	T-AC	T-NAC

4.3 The payoffs

The agents' payoffs vary according to the treatment, whether the technology is actually sold to the consumer and if a damage happens. Whatever the price paid (1400 or 1500 ECUs) by the consumer, the scientist receives 1/3 of the amount and the producer keeps 2/3. Thus the scientist's payoffs, π^S changes according to the situations:

$$\pi^S = \begin{cases} 1000 - \text{Investment} & \text{if technology is not sold} \\ 1000 - \text{Investment} + 1/3 \text{ sale price} & \text{if technology is sold and S is not liable or if no damage happens} \\ 1000 - \text{Investment} + 1/3 \text{ sale price} - 800 & \text{if technology is sold and S is liable for 1/3} \\ 1000 - \text{Investment} + 1/3 \text{ sale price} - 1200 & \text{if technology is sold and S is liable for 1/2} \end{cases}$$

Where Investment is the number of ECUs invested by the Scientist in order to increase the probability of developing a Good technology. Similarly we can distinguish the situations giving different payoffs for the

producer:

$$\pi^P = \begin{cases} 1500 - \text{Cost} & \text{if technology is not sold} \\ 1500 - \text{Cost} + 2/3 \text{ sale price} & \text{if technology is sold and if no damage happens} \\ 1500 - \text{Cost} + 2/3 \text{ sale price} - 800 & \text{if technology is sold and P is liable for 1/3} \\ 1500 - \text{Cost} + 2/3 \text{ sale price} - 1200 & \text{if technology is sold and P is liable for 1/2} \\ 1500 - \text{Cost} + 2/3 \text{ sale price} - 2400 & \text{if technology is sold and P is fully liable} \\ 1500 - \text{Cost} + 2/3 \text{ sale price} - 2400 - 500 & \text{if technology is sold and P is fully liable and if the fine applies} \end{cases}$$

Where Cost is equal to 50 ECUs if the Producer pays to find out the true nature of the technology. Cost is zero otherwise. And for the consumer, we have:

$$\pi^C = \begin{cases} 1500 & \text{if technology is not sold} \\ 1500 - \text{Price} + 1800 & \text{if technology is sold and C is not liable and no damage happens} \\ 1500 - \text{Price} + 1200 & \text{if technology is sold and C is not liable and a damage happens} \\ 1500 - \text{Price} + 1200 - 800 & \text{if technology is sold and C is liable for the damage that happens} \end{cases}$$

All five rounds are similar and follow the same sequence of decisions. At the end of each round, all three subjects know if the technology failed and they also know their payoffs for that round.

4.4 Inequality and risk preferences

In addition to the main game, we elicit participants' risk attitude using the method developed by Eckel and Grossman (2002). In this task, subjects are presented with 5 different gambles and have to select only one of them. Each gamble offers a 50% chance of getting the low payoff and a 50% chance of getting the high payoff. The first gamble (Gamble 1) is a certain gamble (no risk) while the fifth one (Gamble 5) is the riskiest one (highest expected return and highest standard deviation). Risk-averse subjects are expected to select the gambles with the lowest standard deviations (see Appendix A.3).

We also elicit distributional preference types by an Equality Equivalence Test (Kerschbamer, 2015). The test asks the subjects to make ten binary choices between an equal and an unequal allocation, involving an own payoff and a payoff for a randomly matched subject (see Appendix A.3). The choices are broken down into a disadvantageous inequality block of five choices and an advantageous inequality block of five choices. These ten choices, and in particular the row at which the subject switches from the equal to the unequal allocation, allow us to classify all subjects according to their distributional preference types. In particular, we are interested in the subjects that are inequality averse.¹⁸

¹⁸See Balafoutas et al. (2012) and Kerschbamer (2015) for more details on the classification. The method allows to classify the subjects into four categories; altruistic, inequality averse, spiteful, and inequality loving. Note that selfish subjects are a subset

4.5 Predictions

Given the theoretical model developed above and the parameters of the experimental design, we can make several predictions to be tested. This is done by calculating optimal decisions for each of the three agents. As in the theoretical model, equilibria are derived under the assumption of risk-neutrality of agents. It is possible to make predictions about how risk-averse or risk-lover agents behave relatively to risk-neutral ones. We present some related predictions in Appendix A.6 but as it will be clear in the results section, risk preferences play little role in the individual decision making.

Starting with the Consumer, we can derive predictions about the probability to buy the good and the effort to make when the technology is Non-Autonomous (i.e. treatments T-NAP, T-NAS and T-NAC). First, given the experiment's parameters and specifications, a risk-neutral Consumer is expected to buy the good, whatever its quality (G or B), whatever the technology being Autonomous or Non-Autonomous, and whatever the sharing rule of liability. However, the Consumers' benefit from consuming the good decreases with the degree of liability $l_C H$.

Prediction 1

A risk-neutral Consumer always has an interest in buying the good, but the incentive to buy decreases with the level of liability.

Given the experiment's parameters, in all treatments, the Consumer's expected benefit from consuming the goods exceeds the costs including the price to be paid and the expected damages. Let us remind that the (expected) balance is positive for what concerns monetary payoffs but there is also a cost for cognitive efforts when the technology is Non-Autonomous (in T-NAP, T-NAS and T-NAC) that we do not take into account in this calculation. Indeed, when the technology is Non-Autonomous, the Consumer can reduce the probability of causing an accident through the effort task, but the marginal decrease in probabilities is bigger for a B-quality than for a G-quality. Thus, for a given liability level, the marginal benefit of effort is higher with a B-quality than with a G-quality. Moreover, for a given quality of technology, the marginal benefit of effort increases with the amount the Consumer has to pay in damages, $l_C H$. Assuming that the (cognitive) cost function of effort is the same whatever the quality of the technology, it results in the following Prediction.

Prediction 2

1. The Consumer exerts a higher (or equal) level of effort when faced with a B-quality technology, than when faced with a G-quality technology, for a given level of liability.

2. The level of consumer's effort increases with the level of liability.

of the four other categories. We could including them in a separate category but it does not affect the results. In the econometric specification, we include a control for self-interested individual coming from the socio-demographic questionnaire.

It is important to notice that, given that the cognitive effort has a non-monetary cost, we are unable to predict an equilibrium effort. To facilitate numerical calculations, we assume that $p_G(\epsilon_G^*) = 0.15$ and $p_B(\epsilon_B^*) = 0.23$.¹⁹ Let us now turn to the behavior of the Producer. Our theoretical model above shows that in case of an Autonomous technology, liability (alone) provides no incentives for the Producer to pay for information about the quality of the technology and to announce the true quality to the Consumer. This is because the inherent risk is exogenous: neither the Producer nor the Consumer can control it. Thus the parameters of the experiment have been chosen such that we expect the Producer to pay $K = 50$ for obtaining information on the true quality of the technology in case of Non-Autonomous technology but, in case of an Autonomous technology, the Producer should pay for information only when liability is associated with a fine $F_P = 500$ (for failure in declaring a B-quality technology, in treatment T-AF).

In the case of a Non-Autonomous technology, the incentive for the Producer to pay for information comes from the possibility to rightly advice the Consumer and to avoid announcing the wrong quality. Indeed, if the Consumer makes a low level of effort to reduce the probability of accident, thinking that the technology is of Good quality (while it is a B-quality), then the level of risk of accident could be high, which can be costly for the Producer. Let us remind that the Producer always share a part of liability in case of a damage.²⁰ Thus in case of a Non-Autonomous technology, the Producer has higher incentives to pay for information when the degree of liability is higher. This also means that when the Producer pays for information, he uses this information and declares the true quality to the Consumer.

Prediction 3

1. *In case of “Non-Autonomous” technology, the risk-neutral Producer invests in information seeking and declares the true type to the Consumer*
2. *In case of “Non-Autonomous” technology, the incentive for the Producer to invest in information seeking increases with his share of liability, and so he is less prone to lie.*
3. *In case of “Autonomous” technology, the Producer does not pay for information and always declares the technology to be of G-quality to the Consumer, except if he can be fined in case of failure for declaring a B-quality technology (following an accident). In that case, the Producer pays for information and declares the true type.*

Finally, the Scientist first chooses the level of investment e to increase the probability to obtain a G-quality technology and then decides which quality of technology to announce to the Producer. Since a G-quality technology is sold at a higher price than a B-quality, there is always a strictly positive incentive to invest

¹⁹Recall that the Consumers’ effort to reduce the probability of an accident consists in counting the number of 1 in a series of tables composed by 0 and 1. As showed in Table 2, from 0 to 4 rightly counted tables, the probability of a damage decreases. $p_G(\epsilon_G^*) = 0.15$ is obtained for 1 counted table, and $p_B(\epsilon_B^*) = 0.23$ is obtained for 3 counted tables.

²⁰Since the Consumer makes a lower level of effort when facing a G-quality, applying such a low level of effort with a B-quality could lead to a high probability an accident to occur. In such a situation, the Producer faces high expected damages, which gives incentives to pay for information. In our specification, $p_B(\epsilon_G^*) = 0.45$.

for the Scientist. Furthermore, the incentive to invest increases with the difference in expected damages between selling a good endowed with a B-quality technology and selling a good endowed with a G-quality one. For a given level of liability, this difference is higher in the case of an Autonomous than in a case of a Non-Autonomous technology.²¹ However, as shown in the theoretical analysis above, the Scientist always announces a G-quality technology in case of an Autonomous technology. This is because once the quality of the technology is given, the risk is exogenous and cannot be controlled by anyone.

Prediction 4

1. *The Scientist always has incentives to invest in order to reduce the probability of designing a B-quality technology, but incentives are always lower than first-best ones.*
2. *Incentives to invest increase with his share of liability.*
3. *For a given level of liability, incentives to invest are higher in case of Autonomous technology than in case of a Non-Autonomous one.*
4. *In case of Non-Autonomous technology, the Scientist's incentives to declare the true quality of the technology to the Producer increase with his share of liability.*
5. *In case of Autonomous technology, the Scientist always announces a G-quality technology to the Producer.*

Table A.6 in the Appendix A.7 presents a summary of all our predictions and equilibria. This is done for each agent, for all treatments, given our parameters values, specifications and assuming risk-neutrality.

5 Results

5.1 Procedure

A total of 504 subjects participated in 21 sessions (3 sessions per treatment) in October 2022 and September and October 2023 at the Laboratory of Experimental Economics in Strasbourg (LEES). The subjects were recruited from a list of experimental subjects maintained at the LEES using the ORSEE software (Greiner, 2015). The experiment was computerized with the webplatform *EconPlay*²². Upon arrival, each subject was randomly assigned to a computer. The instructions were read aloud by the experimenter and, before starting, a comprehension questionnaire was administered to check that the rules were well understood. All questions were answered privately. Then the main game took place, followed by the elicitation of risk preferences, the elicitation of the social preferences and finally a post-experimental questionnaire.

²¹Recall that in the presence of an Autonomous technology, the probability of accident cannot be reduced by the Consumer. Levels of probabilities are the highest, as well as the difference in probabilities between B-quality and G-quality technologies ($p_B - p_G = 0.6 - 0.2 = 0.4$). So, for a given payment in damages, the difference in expected damages is higher in the case of an Autonomous than in the case of a Non-Autonomous technology.

²²www.econplay.fr

At the end of the experiment, one period from the main game was drawn randomly for actual payment. A random draw was also made to pick the payoff earned by subjects in the risk elicitation task. The conversion rate was 1,000 ECUs to €7.5 for the main game and earnings for the risk aversion elicitation and the social preferences tasks are expressed directly in euros. Subjects were paid their earnings in a separate room and privately at the end of the session. Average earnings were €20 (std. dev. = 4.63). The experiment lasted 70 minutes on average.

In the following subsections, we present the results by looking successively at the three agents decisions. For each one, we present average values and perform a series of non parametric tests. Then we examine the individual choices in econometric specifications wherein we control for individual characteristics in order to identify the effects of the treatments on subjects' behavior.

5.2 The consumer

Figure 1 displays the proportion of consumers who buy the good in each treatment and according to the quality of technology announced by the producer. In each treatment, the probability to buy the good is higher when the producer announces that a G-quality technology is embedded. On average and for each level of sharing liability, consumers buy the good more often when the technology is non autonomous, that is when they can somehow control the risk of a damage²³. Whatever the nature of the technology (autonomous or not), we observe that the consumers buy the good less often when they share liability with the other agents (in T-NAC and T-AC)²⁴, which tends to confirm Prediction 1. Looking at the possibility to reduce the probability of a damage, we see on Figure 2 that within each level of sharing liability, there is no difference according to the quality of technology announced by the producer. This contradicts Prediction 2.1. The average effort appears to be bigger when the consumer is liable (in T-NAC and T-AC) but Mann-Whitney tests taking the individual averages as reference does not show any significant difference between treatments. Looking at the evolution through rounds, Figures A.1 and A.2 show the proportion of consumers who buy the good as well as the effort made in each round. We do not observe important variation along the periods in the willingness to buy the good, whatever the treatment condition. Although, consumers apply a similar effort in the first round in all three non-autonomous treatment, they increase this effort along time but much more when they are liable for the damage (in T-NAC). While the increase observed in the three conditions can be explained by an improvement in their capacity to solve the task (i.e. some kind of learning effect), the differential with T-NAC is perhaps the result of an increased willingness to avoid a damage when one has experienced one in the past. Something we are going to control for in the econometric analysis below.

²³Taking the individual average over the period, Mann-Whitney tests show significant difference between T-NAP and T-AP ($z=2.307, p=0.021$), between T-NAS and T-AS ($z=2.948, p=0.003$) and between T-NAC and T-AC ($p=2.496, z=0.012$).

²⁴Taking the individual average over the period, Mann-Whitney tests show significant difference between T-NAC and T-NAP ($z=1.731, p=0.083$), between T-NAC and T-NAS ($z=3.201, p=0.001$), between T-AC and T-AP ($z=2.287, p=0.022$) and between T-AC and T-AS ($z=3.351, p=0.001$).

Figure 1: Proportion of consumers who buy the technology by treatment and producer's announcement

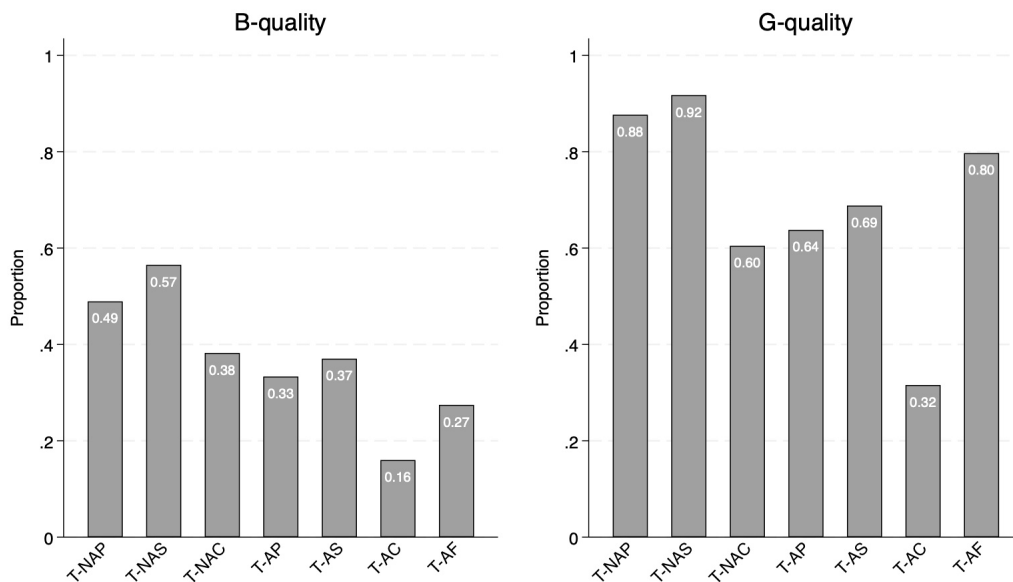
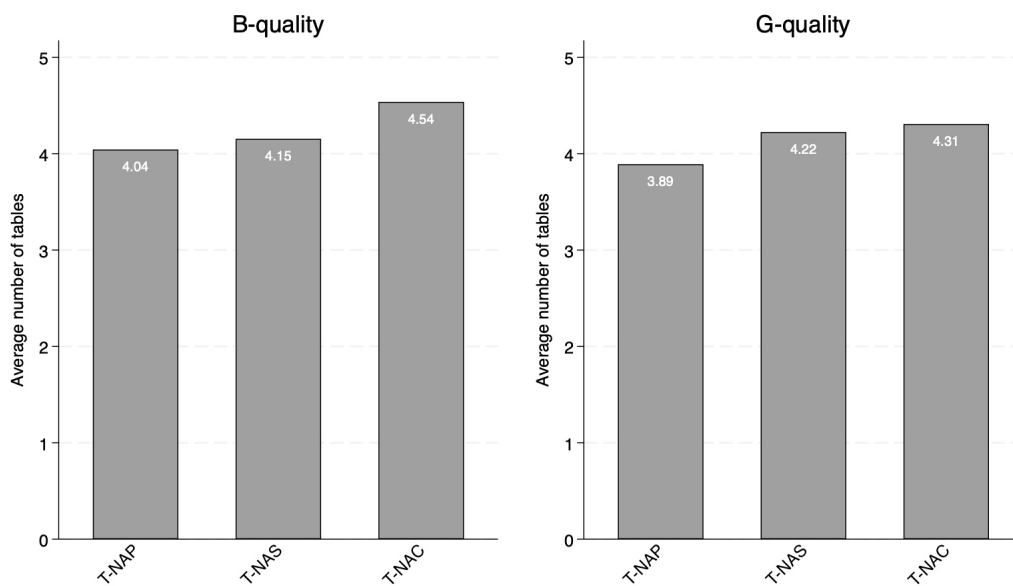


Figure 2: Average consumer's effort by treatment and producer's announcement



These descriptive results do not hold constant individual characteristics and heterogeneity among subjects. To control these factors, we estimate multivariate models in Table 4. In a first step, we try to understand the demand for the good. Specifications (1) to (3) are logit models where the dependent variable is equal to one if the consumer buy the good, zero otherwise. We include a dummy for each treatment and we control for the quality that is announced by the producer as well as individual characteristics such as age, gender, level

and field of study. We also introduce risk preference as measured by the lottery task, if the subject declares to trust in others and if he considers to be self-interested²⁵. We estimate the same regression separately for autonomous and non autonomous technology in specifications (1) and (2) but in specification (3) we look at the isolated effect of the autonomy when we gather all treatments in one regression.

Table 4: Consumers decisions

	Bought the good			Effort
	Non autonomous (1)	Autonomous (2)	All (3)	(4)
T-NAP	Ref			Ref.
T-NAS	0.063 (0.046)			0.431* (0.182)
T-NAC	-0.202*** (0.056)			0.551** (0.201)
T-AP		Ref.		
T-AS		-0.013 (0.070)		
T-AC		-0.287*** (0.061)		
T-AF		0.091 (0.062)		
Autonomous			-0.186*** (0.035)	
Producer announces G-quality	0.295*** (0.040)	0.334*** (0.039)	0.294*** (0.029)	-0.083 (0.227)
Risk-Seeking	0.013 (0.013)	-0.005 (0.015)	0.004 (0.010)	0.089 (0.057)
Self-interested individuals	-0.001 (0.014)	0.033** (0.010)	0.015 (0.008)	0.001 (0.043)
Trust in others	0.113* (0.051)	0.130** (0.046)	0.140*** (0.034)	0.241 (0.211)
Period	0.020 (0.013)	-0.026 (0.014)	-0.007 (0.010)	0.220*** (0.062)
Constant				1.653 (0.883)
Observations	360	480	840	254

Notes: Each regression includes controls for age, gender, whether the subjects is a bachelor student and studying economics or management. Average marginal effects are reported for Logit estimation. Standard errors are clustered at the individual level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

In Table 4, we report only the main findings but complete tables are available in the appendix. Results of specifications (1) and (2) confirm our descriptive findings that sharing liability with the consumer in case of an accident reduce the consumer's propensity to buy the good. When the producer announces a G-quality technology, it increases the probability that the consumer buys the good. Surprisingly, risk preference does not affect the choice to buy²⁶ but trust in other does. We could expect that risk preference plays a role according to the producer's announcement but an interaction between both does not give significant results. We also looked at the effect of having experienced a damage in the past and it does not release significant neither (results are not reported here). In specification (3), we gather all treatments and look at the effect of the autonomy alone, we observe that it reduces the probability to buy the good.

The last specification in Table 4 presents an OLS regression wherein the dependent variable is the effort made

²⁵See Appendix A.2 for the list of socio-demographic questions asked to subjects.

²⁶We have tried different definition of risk seeking with the results of the lottery task and it does change the null result we report here.

by the consumers in case of a non autonomous technology²⁷. The results show a positive and significant effect of sharing liability on the effort made by the consumer which confirms prediction 2.2. As explained above, past experience of a damage could influence the propensity to apply some effort but the introduction of an indicator of a damage in the past rounds or in the last round releases insignificant results without affecting other findings.

Result 1 *Liability increases the consumer's effort to reduce the probability of a damage but it decreases the consumer's propensity to buy the good, whatever its quality or the autonomy.*

5.3 The producer

The producers make two decisions. First, they have to decide to pay for having information about the quality of the technology. Second they have to announce the quality of the technology to the consumer but they can lie about it and announce a G-quality while it is not true.

Figure 3 displays the proportion of producers who pay for obtaining information about the true quality of the technology. This proportion is always positive and relatively high in treatments with autonomous technology. This confirms Prediction 3.1. but prediction 3.3. tells that the producers should not pay for information when facing an autonomous technology. Surprisingly, the proportion is much higher in T-AP without this being explained by some outlying individual choices. In all three sessions we conducted for this treatment condition, we observe about 70% of producers pay. The autonomy of the technology seems to affect producers decisions. This might be explained by the fact that in that situation, they know that the consumer cannot reduce the probability of a damage, and thus he “has to” be aware on the risk while deciding to buy or not the good.

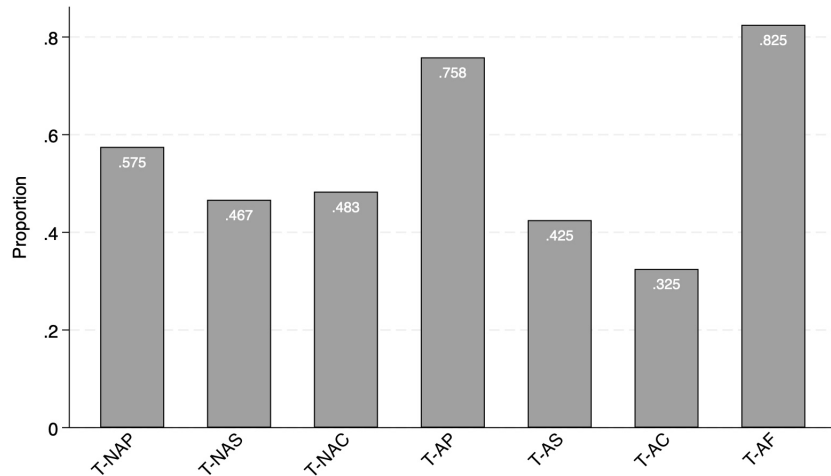
On average, we observe a positive effect of liability on the probability to pay for information. The biggest proportion are observed in treatments T-NAP, T-AP and T-AF when the producer is fully responsible for the damage's cost. Taking the individual average over the periods, a Mann-Whitney test shows a significant difference between treatments where the producers is fully responsible and the others ($z=-5.007$, $p=0.000$). This result is mainly driven by autonomous technology where we observe a clear decrease of proportion when the liability is shared²⁸. These results confirm the role of liability on the probability to pay for information but only in the case of an autonomous technology which contradicts somehow Prediction 3.3. In the appendix, we also present the same figure according to the announcement made by the scientist. We observe the same difference except that the proportion of producers who pay for information is much higher when the announcement is G-quality. Indeed, there is no reason to check for the real type of the technology when the scientist announces a bad one since there is no incentive for him to lie (see below). Thus producers in this

²⁷A Tobit regression gives similar qualitative results.

²⁸Taking the individual average over the periods, Mann-Whitney tests show significant difference between T-AP and T-AS ($z=2.753$, $p=0.006$) and between T-AP and T-AC ($z=4.455$, $p=0.000$).

situation seem to need to be reassured anyway. Still in the appendix we present the evolution across rounds in each treatment. Interestingly, in each treatment condition, except T-AF, we observe a small decline of the propensity to pay along the periods. In T-AF, the fine appears to be a strong incentive against any risk taking and, as we show below, against propensity to lie.

Figure 3: Proportion of producers who pays for information by treatment

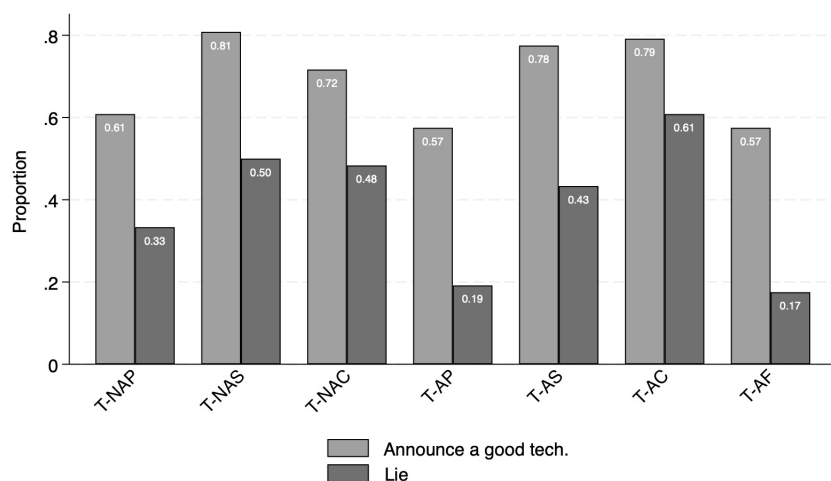


On Figure 4, both the proportion of producers who announce a G-quality technology and the proportion who lie are presented. We do not observe significant difference in the proportion of producers who announce a good technology between autonomous and non autonomous treatments. In particular under autonomous technology (except when the producer can be fined) we expect the producers to always announce a G-quality technology. The same is true when we look at the proportion of producers who lie (that is they announce a G-quality technology while they have been announced a bad technology²⁹). The proportion of lies decreases with the level of sharing liability and this for both autonomous and non autonomous technologies.³⁰ This confirms Prediction 3.3 about non autonomous technology. However in the case of an autonomous technology, the producer should always announce a G-quality (and thus lie) whatever the liability rule.

²⁹We consider here that the producers lie when they deviate from what the scientist announced. We could consider a more restricted definition such that the producers only lie when they did know the very quality of the technology with certainty (namely after having paid for information). It gives similar conclusions.

³⁰Taking the individual average over the periods, a Mann-Whitney test shows a significant difference between T-NAP and T-NAS ($z=-2.211$, $p=0.082$) T-NAP and T-NAC ($z = -2.315$, $p = 0.071$), T-AP and T-AS ($z=-2.357$, $p=0.018$), T-AP and T-AC ($z=-4.075$, $p=0.000$).

Figure 4: Proportion of producers who declare a Good technology and/or lie



Similarly to the consumer, we run a series of regressions in order to control for individual characteristics. All specifications presented in Table 5 are logit models. In specifications (1) to (3), the dependent variable takes the value one if the producer paid for information about the type of technology, zero otherwise. The control variables are the same as in Table 4 but we introduce one additional dummy variable equal to one if the subject is inequality averse, zero otherwise. In specifications (1) and (2), we test the effect of liability on the payment for information according to the type of technology. As for the descriptive results above, there is no effect of liability on the probability to pay for information with non autonomous technology so that we cannot confirm Prediction 3.2. However we observe an important negative impact of sharing liability in the autonomous treatments which contradicts Prediction 3.3.

In specification (3), we look at the effect of the autonomy of the technology. It has an important effect on the probability to pay for information. This may be due to the fact that, in that case, the producer cannot expect the consumer to make effort to reduce the probability of an accident. The producer may then prefer to know exactly the type of technology before announcing one or another. Having the scientist that announces a G-quality technology increases the probability to pay for information in all three specifications. This is expected since there is no reason for the scientist to announce a B-quality technology if it is not the case. As for the descriptive analysis, we observe a diminishing trend over the periods. In additional regressions we include past damage as an explaining factor but it does not show any significant effect.

Both paying for information and lying are intrinsically related because the producer's willingness to obtain information is dependent on the willingness to tell the truth. We look at the decision to lie below but in this experiment, there is no reputation effect that could affect the willingness to lie since subjects are matched each round with a new consumer-scientist pair. However we cannot reject that some subjects have an intrinsic preference for truth-telling as it has been show previously in the literature (see i.e. Ellingsen and Johannesson,

2004; Hurkens and Kartik, 2009; Fischbacher and Föllmi-Heusi, 2013). Also, subjects may refrain from lying because they do not want to inflict losses on other participants, i.e. they care about the payoff of other participants (Gneezy, 2005). Indeed, in Table 5, the effect of inequality aversion appears to be positive and highly significant in specification (2) and (3). In the context of our experiment, this may be explained by the fact that in treatments with autonomous technology, the Producer knows that the Consumer has no possibility to correct the high probability of damage. The producer can then be less willing to cheat and thus want to know the true state of the technology.

Thus in Specifications (4) to (6), we look at the treatment effect on the probability to lie. We observe a strong effect of liability on the probability to lie in both types of technology which confirms Prediction 3.2 but contradicts Prediction 3.3. On the contrary to specifications (1) to (3), no other controls release significant except the rounds that show an increasing trend toward lying. This is interesting since it tends to show that subjects learn about the strategy to play in the game. In specification (6), autonomy appears to have a negative effect on the propensity to lie. As explained above, in that case, the Producer knows that the Consumer cannot control the risk and could be less inclined to cheat.

Table 5: Producers decisions

	Paid for information			Lied		
	Non autonomous (1)	Autonomous (2)	All (3)	Non autonomous (4)	Autonomous (5)	All (6)
T-NAP	Ref.			Ref.		
T-NAS	-0.066 (0.077)			0.183* (0.076)		
T-NAC	-0.010 (0.057)			0.112 (0.059)		
T-AP		Ref.			Ref.	
T-AS		-0.282*** (0.057)			0.257*** (0.057)	
T-AC		-0.373*** (0.068)			0.432*** (0.066)	
T-AF		0.015 (0.049)			-0.039 (0.044)	
Autonomous			0.098** (0.038)			-0.099* (0.039)
Scientist announces G-quality	0.267*** (0.055)	0.353*** (0.038)	0.369*** (0.033)	-0.053 (0.063)	0.037 (0.046)	-0.052 (0.039)
Risk-Seeking	-0.055** (0.018)	0.005 (0.016)	-0.004 (0.013)	0.028 (0.020)	0.016 (0.018)	0.011 (0.014)
Self-interested individuals	-0.003 (0.013)	-0.001 (0.009)	0.007 (0.008)	-0.008 (0.013)	0.016 (0.011)	0.001 (0.008)
Trust in others	-0.005 (0.069)	0.061 (0.044)	0.021 (0.039)	-0.048 (0.067)	-0.023 (0.047)	-0.027 (0.039)
Inequality averse	0.052 (0.060)	0.137*** (0.041)	0.150*** (0.038)	-0.029 (0.062)	0.026 (0.052)	-0.034 (0.039)
Round	-0.055*** (0.016)	-0.050*** (0.011)	-0.054*** (0.009)	0.037* (0.017)	0.059*** (0.013)	0.048*** (0.010)
Observations	360	480	840	360	480	840

Notes: All columns present Logit estimations and report average marginal effects. Each regression includes controls for age, gender, whether the subjects is a bachelor student and studying economics or management. Average marginal effects are reported for Logit estimation. Standard errors are clustered at the individual level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Result 2 *On average, whatever the treatment, producers pay for information about the technology's quality but the propensity to pay increases with the level of liability when the technology is autonomous. Also, producers*

pay more often for information when the scientist announces a G-quality technology.

The proportion of lies decreases with the level of sharing liability and this for both autonomous and non autonomous technologies.

5.4 The scientist

The scientist has to decide how much to invest in order to increase the probability of getting a G-quality technology. Figure 5 shows the average amount invested by the scientists. In each treatment, we observe a positive amount of investment that is different than zero according to t -tests, which confirms prediction 4.1. Also, the investment increases when the scientist becomes liable for possible damages. This is particularly true when the scientist support half of the cost of damage (in treatments T-NAS and T-AS)³¹, which confirms prediction 4.2. In the appendix, Figure A.6 presents the evolution of investment across rounds. We observe a decreasing trend of the investment towards levels lower than predicted equilibrium. Indeed, in all treatments, the scientist’s investment starts at high level an then decreases to low levels, which confirms Prediction 4.1. For a given level of liability, we expect the investment to be higher in Autonomous technology than in Non-Autonomous technology. Figure 5 shows a higher investment in T-AP than in T-NAP ($z=-1.898$, $p=0.057$) but no difference between T-NAS and T-AS ($z=-0.943$, $p=0.346$). Surprisingly the investment in T-NAC is higher than in T-AC. This confirms only partially Prediction 4.3.

Figure 5: Average scientist’s investment

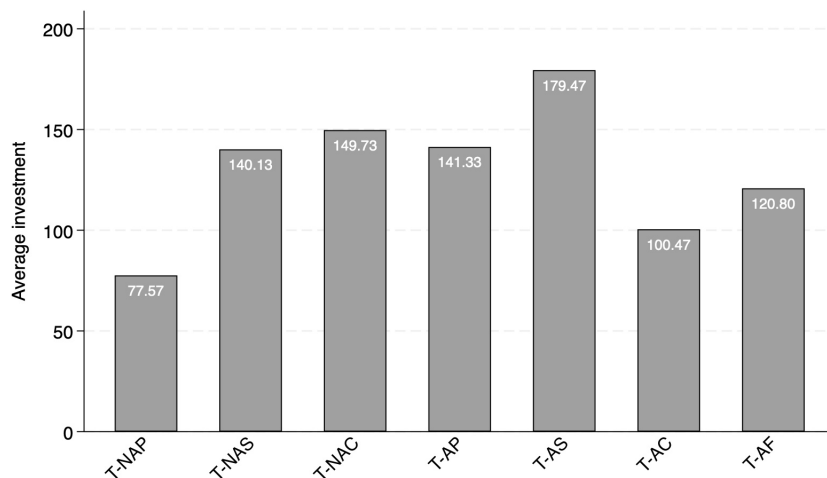


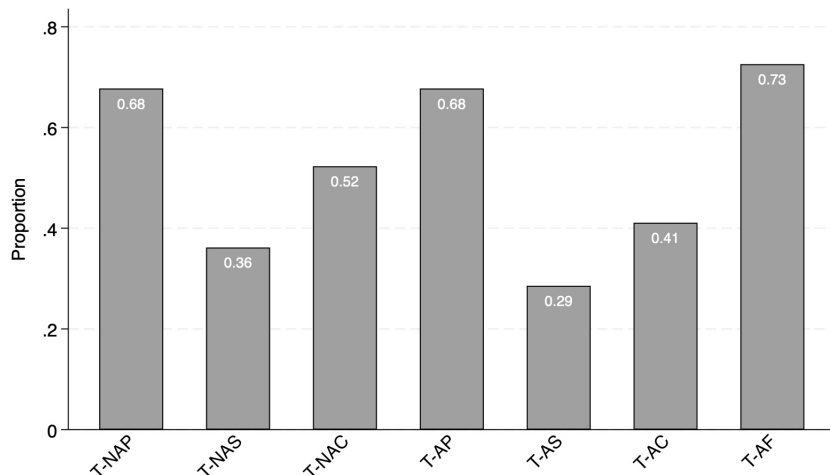
Figure 6 displays the proportion of scientists who lie in each treatment and it appears clearly that the level of sharing liability has an impact on the decision to lie in both types of technology³². These results

³¹Taking the individual average over the periods, Mann-Whitney tests show a significant difference between T-NAP and T-NAS ($z=-2.111$, $p=0.027$), between T-NAP and T-NAC ($z=-2.213$, $p=0.021$) but no significant difference between T-AP and T-AS.

³²Taking the individual average over the periods, Mann-Whitney tests show a significant difference between T-NAP and

confirm prediction 4.4 but we do not observe significant difference between autonomous and non autonomous treatments, which contradicts prediction 4.5.

Figure 6: Proportion of scientists who lie



Turning to econometric analysis, Table 6 presents a series of regressions. We first look at the decision to invest to increase the probability of having a G-quality technology. Specifications (1) to (3) displays OLS regressions where the dependent variable is the amount invested. The decision to invest appears to be driven the level of sharing liability in specifications (1) but not in case of autonomous technology in specification (2). The autonomy of the technology does not affect the the amount invested. This goes in line with predictions 4.1 and 4.2. Trust in other is an important driver of choice in autonomous technology which can be explained that in that case, there is no possibility for the consumer to control the risk. Specifications (4) to (6) present logit models where the dependent variable is the decision to lie or not. The effect of sharing liability is strongly related with a higher probability to lie since we observe a negative and significant effect when the level of liability for the scientist is the highest, whatever the type of the technology. These results confirm also prediction 4.4. However Autonomy of the technology has no effect while theoretical predictions show that, in that case, the scientists should always announce G-quality and then lie (see Prediction 4.5)

Result 3 *On average, the scientist invest a positive amount to improve the technology’s quality but this amount decreases along time towards levels lower than predicted equilibrium.*

Liability decreases the proportion of lies in both types of technology but there is no difference between autonomous and non autonomous technologies.

T-NAS ($z=2.709$, $p=0.007$), between T-NAS and T-NAC ($z=-2.145$, $p=0.032$), between T-AP and T-AS ($z=3.776$, $p=0.000$), between T-AS and T-AC ($z=-2.538$, $p=0.011$)

Table 6: Scientists decisions

	Investment			Lie		
	Non autonomous (1)	Autonomous (2)	All (3)	Non autonomous (4)	Autonomous (5)	All (6)
T-NAP	Ref.			Ref.		
T-NAS	53.141* (24.570)			-0.229*** (0.055)		
T-NAC	53.171* (24.075)			-0.046 (0.063)		
T-AP		Ref.			Ref.	
T-AS		24.552 (25.609)			-0.231*** (0.058)	
T-AC		-45.136 (25.297)			-0.075 (0.065)	
T-AF		-10.482 (25.899)			0.014 (0.068)	
Autonomy of the technology			19.187 (13.372)			0.005 (0.032)
Risk-Seeking	-15.860* (7.593)	2.073 (7.288)	-4.679 (5.150)	0.025 (0.018)	-0.040* (0.016)	-0.003 (0.012)
Self-interested individuals	5.308 (5.557)	-9.494* (4.459)	-4.025 (3.412)	0.011 (0.013)	0.008 (0.011)	0.007 (0.008)
Trust in others	20.357 (26.234)	-84.857*** (20.464)	-42.614** (15.731)	-0.107 (0.063)	0.190*** (0.046)	0.091* (0.036)
Inequality averse	-4.695 (26.593)	-68.084** (22.904)	-29.912 (16.800)	-0.002 (0.063)	0.073 (0.050)	0.061 (0.039)
Period	-17.117* (6.713)	-32.112*** (5.813)	-25.686*** (4.499)	0.000 (0.016)	0.002 (0.014)	0.001 (0.011)
Constant	379.709** (115.216)	117.749 (67.872)	133.699** (51.170)			
Observations	840	360	480	840	360	480

Notes: All columns present Logit estimations and report average marginal effects. Each regression includes controls for age, gender, whether the subjects is a bachelor student and studying economics or management as well as a control for the round of the game. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

5.5 Welfare analysis

Finally, we perform a welfare analysis to compare the effect of the three liability sharing rules in both cases of Autonomous and Non-Autonomous goods (vehicles). To do so, we calculate the social welfare as obtained by summing up all benefits and costs earned and incurred by the three agents and the passive victim. We make the distinction between a situation wherein the social welfare would be reached by a benevolent and omnipotent social planner who would make all decisions (i.e., the first-best), from cases where the social welfare is reached under decentralized policies, i.e., when decisions are made by private agents who are subject to public policy (here, the different ways of sharing liability). The relative desirability of the different liability rules is assessed by comparing their (decentralized) welfare to the welfare reached in the first-best situation.

The social welfare of the first-best situation is computed from equations (1) and (3.3), for Autonomous and Non-Autonomous goods respectively. By applying specifications used to set up the experiment (see sections 4.1 to 4.3, and also section A.5), we can derive first-best values of all decision variables and then calculate the first-best levels of social welfare (both under Autonomous and Non-Autonomous cases).

The social welfare of decentralized policies is obtained in assessing the social impact of decisions which

are taken by the private agents, under each liability sharing rules (given the presence of Autonomous or Non-Autonomous cars). In the presence of Non-Autonomous goods, the social welfare derived from private decisions, $SW_{private}$, is given by

$$\begin{aligned}
SW_{private} = & W_C + W_P + W_S - c(e) + p(e) [(1 - \Phi_S) ((1 - \Phi_P)\Gamma_{C,B}R_{BB} + \Phi_P\Gamma_{C,G}R_{BG}) \\
& + \Phi_S (\Psi(1 - \Phi_P)\Gamma_{C,B}R_{BBI} + \Psi\Phi_P\Gamma_{C,G}R_{BGI} + (1 - \Psi)\Gamma_{C,G}R_{BG}) \\
& + (1 - p(e)) [\Psi\Gamma_{C,G}R_{GGI} + (1 - \Psi)\Gamma_{C,G}R_{GG}]
\end{aligned} \tag{18}$$

with:

$$\begin{aligned}
R_{BB} &= B_{sup} - p_B(e_B^*) (H + (B_{sup} - B_{inf})) \\
R_{BBI} &= B_{sup} - K - p_B(e_B^*) (H + (B_{sup} - B_{inf})) \\
R_{CG} &= B_{sup} - p_G(e_G^*) (H + (B_{sup} - B_{inf})) \\
R_{GGI} &= B_{sup} - K - p_G(e_G^*) (H + (B_{sup} - B_{inf})) \\
R_{BG} &= B_{sup} - p_B(e_G^*) (H + (B_{sup} - B_{inf})) \\
R_{BGI} &= B_{sup} - K - p_B(e_G^*) (H + (B_{sup} - B_{inf}))
\end{aligned}$$

$\Gamma_{C,j}$ is the Consumers' mean buying rate when facing a good which is announced to be j -quality ($j = B, G$), Φ_S is the Scientists' mean rate of lie, Φ_P is the Producers' mean rate of lie and Ψ is the Producer's mean rate of searching for information (when the Scientist announces the technology to be of G-quality). In the Autonomous case, the same rationale applies but probabilities of accidents do not depend on Consumer's behavior and are given by p_B and p_G in case of B-quality and G-quality technologies respectively.

Equilibrium values and first-best values of all decision variables are summarized in Table A.4 for the case of Autonomous cars, and in Table A.5 for the case of Non-Autonomous cars (see in Appendix A.2). By applying these values in the functions defined above, we can calculate the levels of social welfare for the first-best case, and the levels of social welfare for each liability sharing rule, both in the case of Autonomous and Non-Autonomous goods. They are presented in Table 7.

Remind that Autonomous and Non-Autonomous cannot be directly compared each other. In the Autonomous case, the probabilities of causing harm are exogenous. But according to the experimental results above, in both Autonomous and Non-Autonomous cases, a full liability on the Producer is the sharing of liability which provides the highest level of social welfare. From Table A.4 and Table A.5 in Appendix A.2, we can see how (and to what extent) private decisions diverge from the first-best ones. We see that the liability rules which provide the highest levels of social welfare are those which provide the highest buying rates of the good by the

Table 7: social welfare for first-best and treatments (in parentheses)

Technology	Autonomous	Non-autonomous
Liability		
Firs-best	4592	5190
P (Total)	4134,2 <small>(T-AP)</small>	5085,2 <small>(T-NAP)</small>
P (Total) + fine	4199,5 <small>(T-AF)</small>	
P (1/2) and S (1/2)	4132,5 <small>(T-AS)</small>	5082,8 <small>(T-NAS)</small>
P (1/3), S (1/3) and C (1/3)	4066,1 <small>(T-AC)</small>	4652,1 <small>(T-NAC)</small>

Consumers. Indeed, in our setup, buying (and using) the good/car increases the Consumer’s utility, which is source of welfare³³. It is thus of great importance for the Consumer to be confident in the level of safety of the good in order to pay for it. The liability rules that provide the highest buying rates are thus those that either give liability to the Producer, or to the Producer and the Scientist. On the contrary, giving liability to the Consumer sharply decreases the Consumer’s propensity to buy the good - especially in the Autonomous case - which, in turn, decreases the level of social welfare.

It can be underlined that, in the Autonomous case, the possibility for the Producer to be fined provides additional confidence to the Consumer, ensuring the highest buying rate. This, in turn, provides the highest level of social welfare, despite the fact that the fine is also a source of social losses. The fine is a social loss and it provides the Producer with the highest incentives to invest in information acquisition, which is also source of social loss. However, the effect of the fine on the Consumer’s level of confidence (and thus higher buying rate) offsets these social losses.

Result 4 *The highest levels of social welfare are reached when the Producer has maximum liability, and the Consumer minimum liability (no liability).*

In the case of Autonomous goods, the possibility for the Producer to be fined provides additional confidence to the Consumer which increases the Consumer’s buying rate and is thus welfare improving.

6 Conclusion

In this paper, we propose a new theoretical framework to analyze the incentives provided by different allocations of liability in the case of (semi)autonomous devices which are a source of risk of accident. We consider three key agents, an AI provider (scientist), a producer and a consumer, and look at the effect of different rules of sharing liability on the decision making of each type of agent. In addition to the theoretical analysis we test the predictions in an original lab experiment in which we conduct different treatments according to the degree of liability and the autonomy of the good.

Our main theoretical predictions are that a higher level of liability should reduce misbehavior by producers

³³Recall that we suppose that using a good embedded with a technology of B-quality is socially desirable.

and scientist but the degree of incentives of the liability rule depends on the autonomy of the good. Liability should induce higher consumer's effort to reduce the risk of an accident but it should also decrease the incentive to buy the good.

The experimental results confirm some of our predictions. We find that liability is efficient in reducing the proportion of lies by the scientist and the producer. It also increases the consumer's effort to reduce the probability of causing harm but it decreases the consumer's propensity to buy the good, whatever its quality or the autonomy. However, we observe that consumers are less prone to buy a good when they have no control on it (full autonomy). Finally, we find that the scientist invests a positive amount to improve the technology's quality but this amount decreases along time towards levels lower than predicted equilibrium.

We complete our study by making a social welfare analysis. It highlights the importance of letting the producer liable in order to provide the consumer with confidence in the new technology, especially in the case of a full autonomy of the good. This point underlines the need, for the consumer, to be protected by the liability system to ensure the new technology to be adopted.

Bibliography

- Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, 87(4):1115–1153.
- Balafoutas, L., Beck, A., Kerschbamer, R., and Sutter, M. (2013). What drives taxi drivers? a field experiment on fraud in a market for credence goods. *Review of Economic Studies*, 80(3):876–891.
- Balafoutas, L., Kerschbamer, R., and Sutter, M. (2012). Distributional preferences and competitive behavior. *Journal of Economic Behavior & Organization*, 83(1):125–135.
- Beck, A., Kerschbamer, R., Qiu, J., and Sutter, M. (2014). Car mechanics in the lab—Investigating the behavior of real experts on experimental markets for credence goods. *Journal of Economic Behavior & Organization*, 108(C):166–173.
- Briquet, L., Jacob, J., and Lambert, E.-A. (2024). Intelligence artificielle et responsabilité juridique. *mimeo BETA*.
- Brown, J. (1973). Towards an economic theory of liability. *The Journal of Legal Studies*, 2:323–349.
- Brown, R. D. (2021). Property ownership and the legal personhood of artificial intelligence. *Information & Communications Technology Law*, 30(2):208–234.
- Calabresi, G. (1970). *The Cost of Accidents, a Legal and Economic Analysis*. Yale University Press.
- Charness, G. and Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, 74(6):1579–1601.
- Chesterman, S. (2020). Artificial intelligence and the limits of legal personality. *International & Comparative Law Quarterly*, 69(4):819–844.
- Commission, E. (2020). Report on the safety and liability implications of artificial intelligence, the internet of things and robotics. (*COM/2020/64 final*), pages <https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:52020DC0064> (accessed 26 October 2023).
- De Chiara, A., Elizalde, I., Manna, E., and Segura-Moreiras, A. (2021). Car accidents in the age of robots. *International Review of Law and Economics*, 68:<https://doi.org/10.1016/j.irl.2021.106022>.
- Duffy, S. and Hopkins, J. (2013). Sit, stay, drive: The future of autonomous car liability. *SMU Science and Technology Law Review*, 16:453.
- Dufwenberg, M. and Dufwenberg, M. A. (2018). Lies in disguise – a theoretical analysis of cheating. *Journal of Economic Theory*, 175(C):248–264.
- Eckel, C. C. and Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior*, 23(4):281–295.
- Elish, M. and Hwang, T. (2015). Praise the machine! punish the human! *Intelligence and Autonomy Initiative*, page working paper 1.

- Ellingsen, T. and Johannesson, M. (2004). Promises, Threats and Fairness. *Economic Journal*, 114(495):397–420.
- Feltovich, N. (2019). The interaction between competition and unethical behaviour. *Experimental Economics*, 22(1):101–130.
- Fischbacher, U. and Föllmi-Heusi, F. (2013). Lies In Disguise—An Experimental Study On Cheating. *Journal of the European Economic Association*, 11(3):525–547.
- Gless, S., Silverman, E., and Weigend, T. (2016). If robots cause harm, who is to blame? self-driving cars and criminal liability. *New Criminal Law Review*, 19:412–436.
- Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 95(1):384–394.
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1):114–125.
- Guerra, A., Parisi, F., and D., P. (2022a). Liability for robots ii : an economic analysis. *Journal of Institutional Economics*, 18:553–568.
- Guerra, A., Parisi, F., and Pi, D. (2022b). Liability for robots i: legal challenges. *Journal of Institutional Economics*, 18:331–343.
- Hilgendorf, E. (2014). Strafrecht für autos? *Süddeutsche Zeitung*, pages <https://www.sueddeutsche.de/auto/autonomes-fahren-strafrecht-fuer-autos-1.1941244>.
- Hurkens, S. and Kartik, N. (2009). Would i lie to you? on social preferences and lying aversion. *Experimental Economics*, 12(2):180–192.
- Ilkova, V. and A. Ilka, A. (2017). Legal aspects of autonomous vehicles – an overview. *Proceedings of the 2017 21st International Conference on Process Control (PC)*, pages Štrbské Pleso, Slovakia.
- Janal, R. (2016). Die deliktische haftung beim einsatz von robotern – lehren aus der haftung für sachen und gehilfen. *in: Intelligente Agenten und das Recht, Nomos, Baden-Baden, Germany*, 18:141–161.
- Kajackaite, A. and Gneezy, U. (2017). Incentives and cheating. *Games and Economic Behavior*, 102(C):433–444.
- Kalra, N., Anderson, J., and Wachs, W. (2009). Liability and regulation of autonomous vehicle technologies. *California PATH Research Report*, 2009-8.
- Kartik, N. (2009). Strategic communication with lying costs. *Review of Economic Studies*, 76(4):1359–1395.
- Kartik, N., Tercieux, O., and Holden, R. (2014). Simple mechanisms and preferences for honesty. *Games and Economic Behavior*, 83(C):284–290.
- Kelley, R., Schaerer, E., Gomez, M., and Nicolescu, M. (2010).]: Liability in robotics: An international perspective on robots as animals. *Advanced Robotics*, 24:1861–1871.
- Kerschbamer, R. (2015). The geometry of distributional preferences and a non-parametric identification approach: The Equality Equivalence Test. *European Economic Review*, 76(C):85–103.

- Khalmetski, K. and Sliwka, D. (2019). Disguising Lies—Image Concerns and Partial Lying in Cheating Games. *American Economic Journal: Microeconomics*, 11(4):79–110.
- Lopucki, L. M. (2017). Algorithmic entities. *Washington Law Review*, 95:887.
- Matsushima, H. (2008). Role of honesty in full implementation. *Journal of Economic Theory*, 139(1):353–359.
- Mimra, W., Rasch, A., and Waibel, C. (2016). Price competition and reputation in credence goods markets: Experimental evidence. *Games and Economic Behavior*, 100(C):337–352.
- Nevejans, N. (2016). Règles européennes de droit civil en robotique. *Parlement Européen - Direction Générale des Politiques Internes - Affaires Juridiques - Etude PE 571.379*.
- Rothenberg, D. (2016). Can siri 10.0 buy your home? the legal and policy based implications of artificial intelligent robots owning real property. *Washington Journal of Law, Technology & Arts*, 11(5):439–460.
- Schaerer, E., Kelley, R., and Nicolescu, M. (2009). Robots as animals: A framework for liability and responsibility in human-robot interactions. *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pages 72–77.
- Scherer, M. (2016). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. *Harvard Journal of Law & Technology*, 29:353–400.
- Scherer, M. (2018). Of wild beasts and digital analogues: The legal status of autonomous systems. *Nevada Law Journal*, 19:259.
- Shavell, S. (1980). Strict liability versus negligence. *The Journal of Legal Studies*, 9(1):1–25.
- Shavell, S. (2020). On the redesign of accident liability for the world of autonomous vehicles. *The Journal of Legal Studies*, 49:2.
- Solum, L. B. (1992). Legal personhood for artificial intelligences. *North Carolina Law Review*, 70(4):1231–1287.
- Tai, E. (2018). Liability for (semi) autonomous systems: Robots and algorithms. *Research Handbook in Data Science and Law*, Edward Elgar Publishing, pages 55–82.
- Talley, E. (2019). Automatorts: how should accident law adapt autonomous vehicles? lessons from law and economics. *Hoover IP²*, page Working Papers Series No. 19002.
- Tergiman, C. and Villeval, M. C. (2022). The Way People Lie in Markets: Detectable vs. Deniable Lies. *Management Science*, 69(6):3157–3758.
- Vladeck, D. (2014). Machines without principals: Liability rules and artificial intelligence. *Washington Law Review*, 89:117.

A Appendix

A.1 Additional figures

Figure A.1: Proportion of consumers who bought the technology by treatment and period

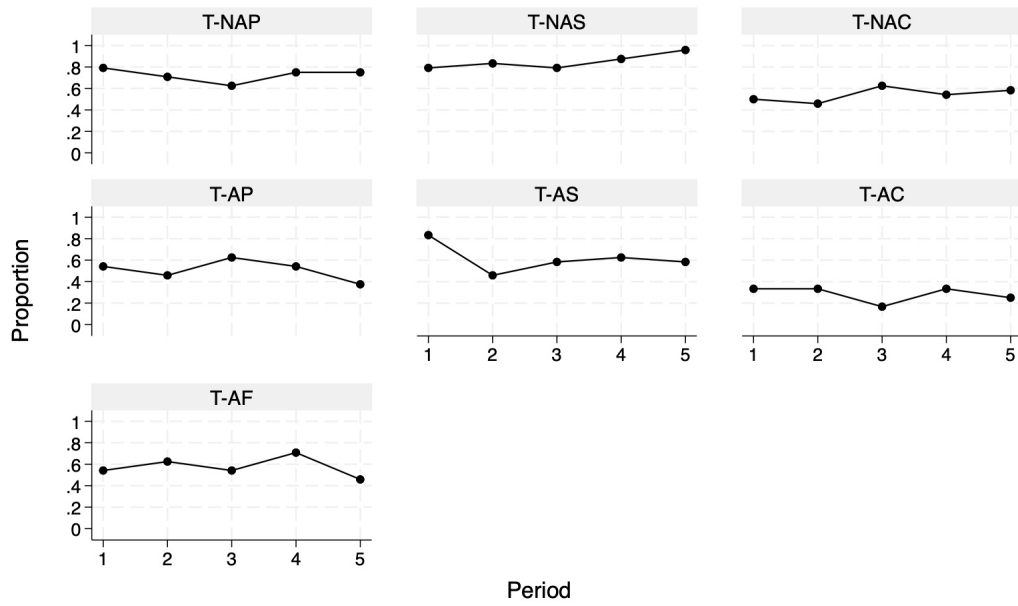


Figure A.2: Number of tables completed by treatment and period

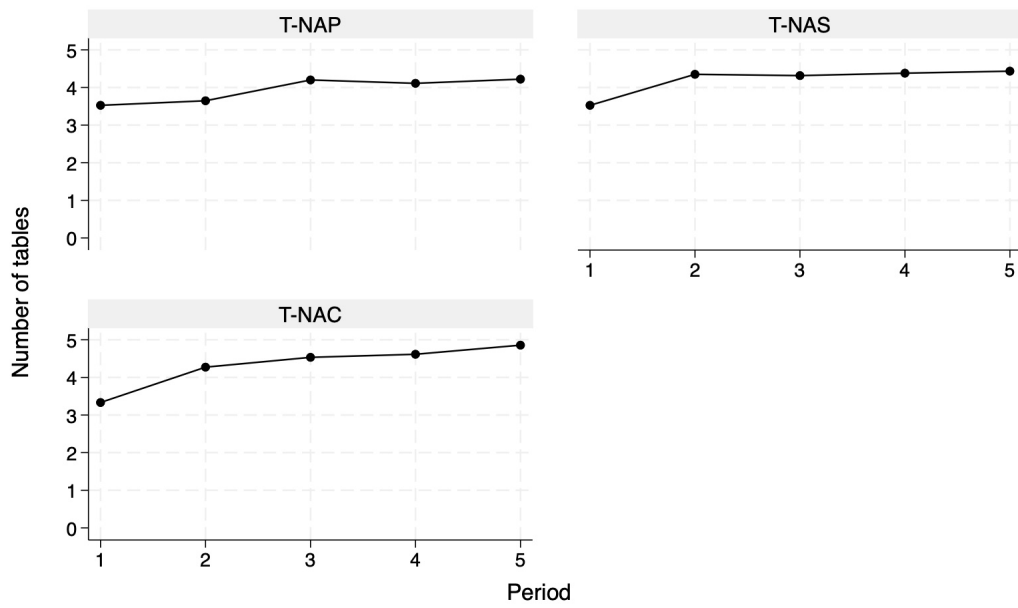


Figure A.3: Proportion of producers who pay for information by treatment and scientist's announcement

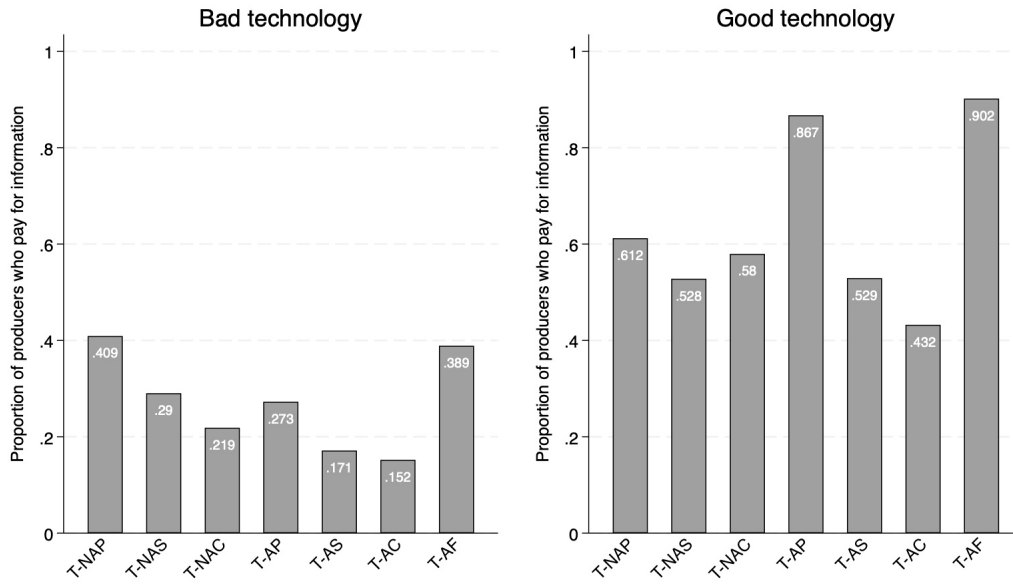


Figure A.4: Proportion of producers who pay for information by treatment and round

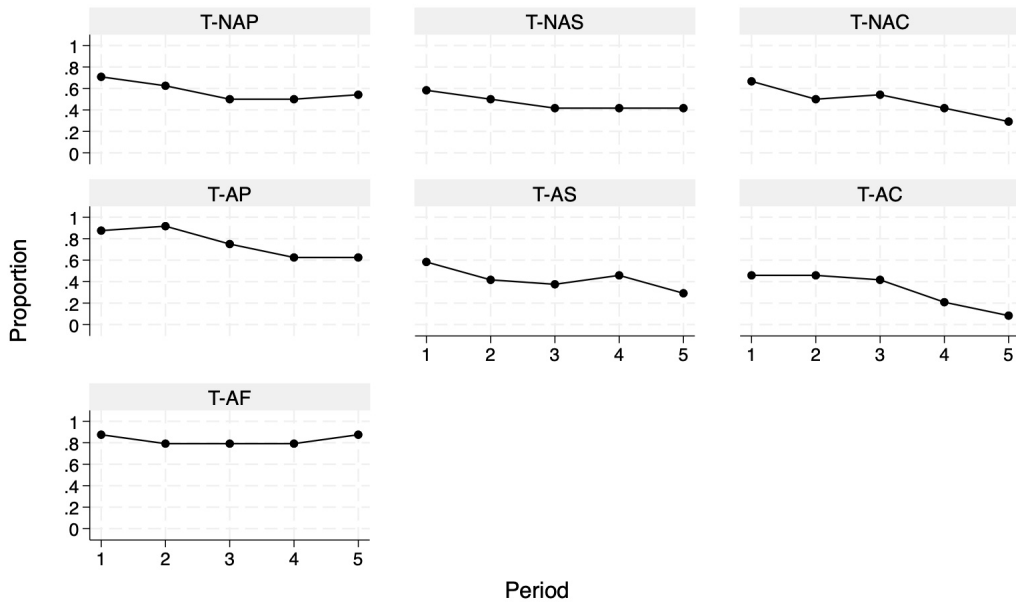


Figure A.5: Proportion of producers who lie by treatment and round

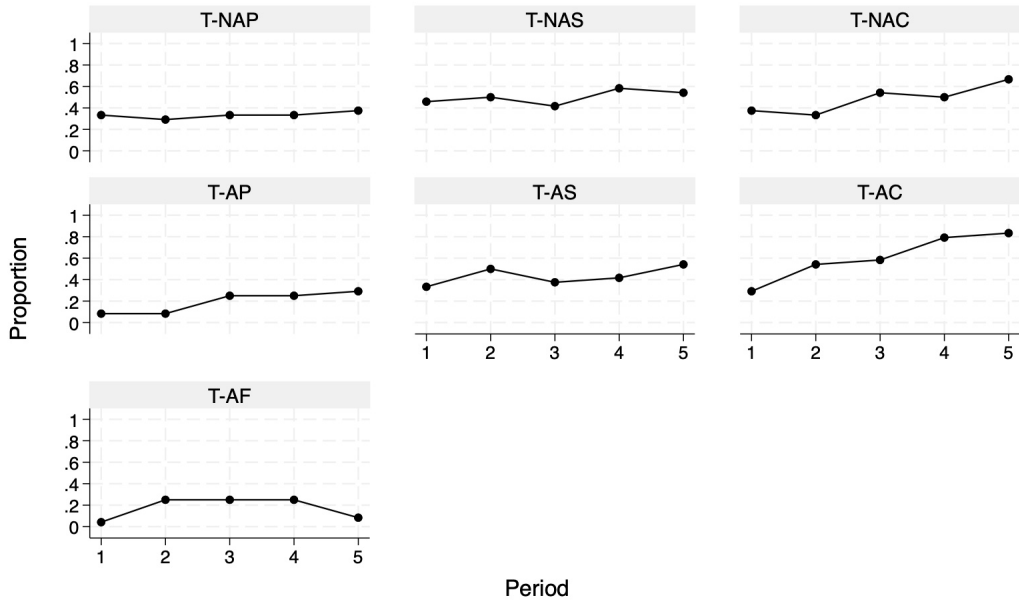


Figure A.6: Average scientist's investment by treatment and rounds

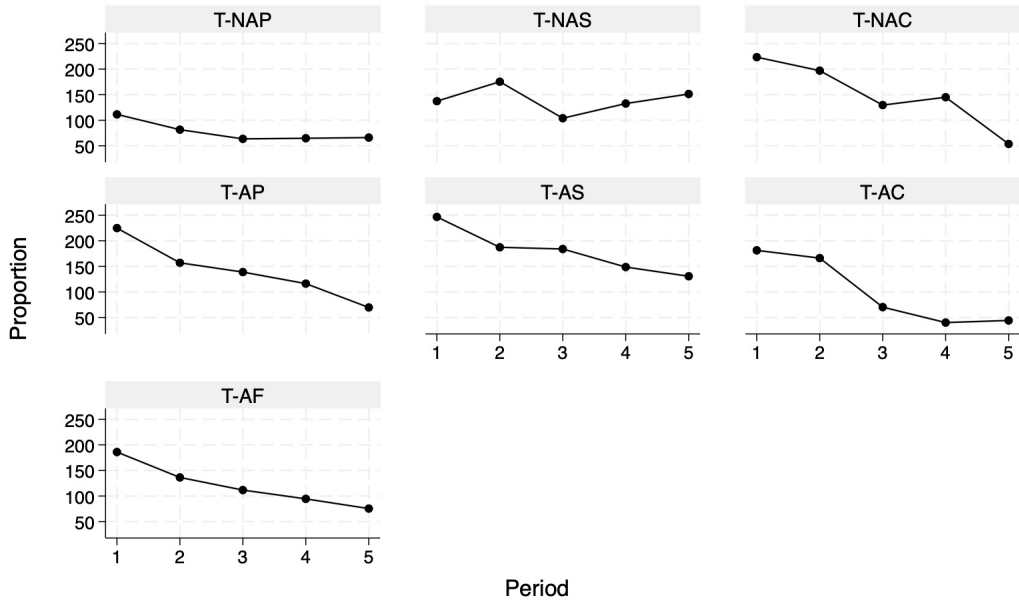
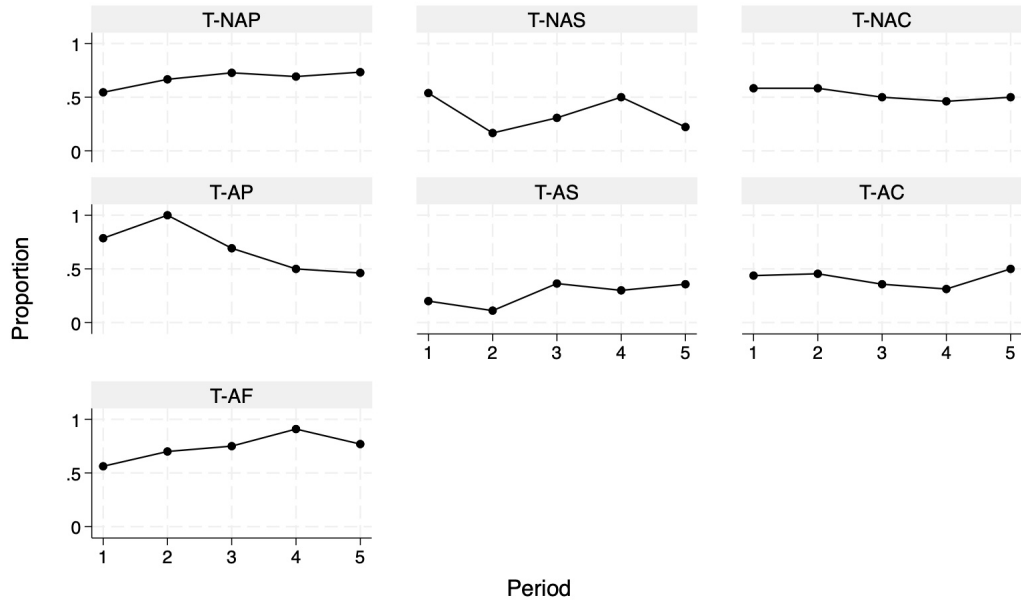


Figure A.7: Proportion of scientists who lie by treatment and rounds



A.2 Additional tables

Table A.1: Consumers decisions

	Bought the Technology			Effort
	Non autonomous	Autonomous	All	
	(1)	(2)	(3)	(4)
T-NAP	Ref			Ref.
T-NAS	0.063 (0.046)			0.431* (0.182)
T-NAC	-0.202*** (0.056)			0.551** (0.201)
T-AP		Ref.		
T-AS		-0.013 (0.070)		
T-AC		-0.287*** (0.061)		
T-AF		0.091 (0.062)		
Autonomous			-0.186*** (0.035)	
Producer announces G-quality	0.295*** (0.040)	0.334*** (0.039)	0.294*** (0.029)	-0.083 (0.227)
Risk-Seeking	0.013 (0.013)	-0.005 (0.015)	0.004 (0.010)	0.089 (0.057)
Self-interested individuals	-0.001 (0.014)	0.033** (0.010)	0.015 (0.008)	0.001 (0.043)
Trust in others	-0.113* (0.051)	-0.130** (0.046)	-0.140*** (0.034)	-0.241 (0.211)
Round	0.020 (0.013)	-0.026 (0.014)	-0.007 (0.010)	0.220*** (0.062)
Age	-0.003 (0.007)	0.016* (0.007)	0.012* (0.005)	0.045 (0.028)
Women	0.010 (0.048)	0.024 (0.047)	0.039 (0.033)	0.362* (0.179)
Eco-Management	0.065 (0.048)	0.013 (0.054)	0.051 (0.037)	0.186 (0.163)
Bachelor	-0.040 (0.042)	0.062 (0.044)	0.029 (0.032)	0.431* (0.195)
Constant				1.653 (0.883)
Observations	360	480	840	254

Notes: Each regression includes controls for age, gender, whether the subjects is a bachelor student and studying economics or management as well as a control for the round of the game. Average marginal effects are reported for Logit estimation. Standard errors are clustered at the individual level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.2: Producers decisions

	Paid for information			Lied		
	Non autonomous	Autonomous	All	Non autonomous	Autonomous	All
	(1)	(2)	(3)	(4)	(5)	(6)
T-NAP	Ref.			Ref.		
T-NAS	-0.066 (0.077)			0.183* (0.076)		
T-NAC	-0.010 (0.057)			0.112 (0.059)		
T-AP		Ref.			Ref.	
T-AS		-0.282*** (0.057)			0.257*** (0.057)	
T-AC		-0.373*** (0.068)			0.432*** (0.066)	
T-AF		0.015 (0.049)			-0.039 (0.044)	
Autonomy of the technology			0.098** (0.038)			-0.099* (0.039)
Scientist announces G-quality	0.267*** (0.055)	0.353*** (0.038)	0.369*** (0.033)	-0.053 (0.063)	0.037 (0.046)	-0.052 (0.039)
Risk-Seeking	-0.055** (0.018)	0.005 (0.016)	-0.004 (0.013)	0.028 (0.020)	0.016 (0.018)	0.011 (0.014)
Self-interested individuals	-0.003 (0.013)	-0.001 (0.009)	0.007 (0.008)	-0.008 (0.013)	0.016 (0.011)	0.001 (0.008)
Trust in others	-0.005 (0.069)	0.061 (0.044)	0.021 (0.039)	-0.048 (0.067)	-0.023 (0.047)	-0.027 (0.039)
Inequality averse	0.052 (0.060)	0.137*** (0.041)	0.150*** (0.038)	-0.029 (0.062)	0.026 (0.052)	-0.034 (0.039)
Round	-0.055*** (0.016)	-0.050*** (0.011)	-0.054*** (0.009)	0.037* (0.017)	0.059*** (0.013)	0.048*** (0.010)
Age	-0.025** (0.009)	0.015* (0.007)	-0.002 (0.004)	0.011 (0.006)	-0.012 (0.008)	0.002 (0.004)
Women	0.082 (0.045)	0.066 (0.038)	0.084** (0.030)	-0.129** (0.048)	-0.000 (0.041)	-0.058 (0.032)
Eco-Management	0.024 (0.052)	0.007 (0.053)	0.022 (0.035)	0.033 (0.059)	-0.016 (0.059)	-0.025 (0.041)
Bachelor	0.113* (0.048)	0.070 (0.036)	0.030 (0.031)	-0.092 (0.051)	-0.094* (0.044)	-0.049 (0.034)
Observations	360	480	840	360	480	840

Notes: All columns present Logit estimations and report average marginal effects. Each regression includes controls for age, gender, whether the subjects is a bachelor student and studying economics or management. Average marginal effects are reported for Logit estimation.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.3: Scientists decisions

	Investment			Lie		
	All	Non autonomous	Autonomous	All	Non autonomous	Autonomous
	(1)	(2)	(3)	(4)	(5)	(6)
T-NAP	Ref.			Ref.		
T-NAS	53.141*			-0.229***		
	(24.570)			(0.055)		
T-NAC	53.171*			-0.046		
	(24.075)			(0.063)		
T-AP		Ref.			Ref.	
T-AS		24.552			-0.231***	
		(25.609)			(0.058)	
T-AC		-45.136			-0.075	
		(25.297)			(0.065)	
T-AF		-10.482			0.014	
		(25.899)			(0.068)	
Autonomy of the technology			19.187			0.005
			(13.372)			(0.032)
Risk-Seeking	-15.860*	2.073	-4.679	0.025	-0.040*	-0.003
	(7.593)	(7.288)	(5.150)	(0.018)	(0.016)	(0.012)
Self-interested individuals	5.308	-9.494*	-4.025	0.011	0.008	0.007
	(5.557)	(4.459)	(3.412)	(0.013)	(0.011)	(0.008)
Trust in others	-20.357	84.857***	42.614**	0.107	-0.190***	-0.091*
	(26.234)	(20.464)	(15.731)	(0.063)	(0.046)	(0.036)
Inequality averse	-4.695	-68.084**	-29.912	-0.002	0.073	0.061
	(26.593)	(22.904)	(16.800)	(0.063)	(0.050)	(0.039)
Period	-17.117*	-32.112***	-25.686***	0.000	0.002	0.001
	(6.713)	(5.813)	(4.499)	(0.016)	(0.014)	(0.011)
Age	-11.201*	4.742*	2.570	0.015	0.008	0.007
	(4.641)	(2.098)	(1.885)	(0.010)	(0.005)	(0.004)
Women	29.368	24.037	29.490*	-0.036	0.033	0.017
	(20.411)	(18.267)	(13.600)	(0.048)	(0.042)	(0.033)
Eco-Management	15.758	7.248	-15.577	0.002	0.004	0.021
	(26.852)	(21.412)	(15.894)	(0.062)	(0.050)	(0.037)
Bachelor	15.146	17.699	22.664	0.025	0.009	-0.006
	(20.144)	(17.815)	(13.195)	(0.049)	(0.042)	(0.032)
Constant	379.709**	117.749	133.699**			
	(115.216)	(67.872)	(51.170)			
Observations	360	480	840	360	480	840

Notes: All columns present Logit estimations and report average marginal effects. Each regression includes controls for age, gender, whether the subjects is a bachelor student and studying economics or management as well as a control for the round of the game. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.4: Mean observed equilibrium values and first-best values, Autonomous case

Treatments	Agents	Scientist		Producer		Consumer	
		Effort	Lie (rate)	Info. (rate)	Lie (rate)	Buy if G (rate)	Buy if B (rate)
P (Total) - TAP		4,03	0,68	0,758	0,19	0,64	0,33
P (1/2) and S (1/2) - TAS		4,13	0,29	0,425	0,43	0,69	0,37
P (1/3) and S (1/3) and C (1/3*) - TAC		3,57	0,41	0,325	0,61	0,32	0,16
P (Total) + fine - TAF		3,89	0,73	0,825	0,17	0,8	0,27
<i>First-best</i>		4	0	0	0	1	1

Table A.5: Mean observed equilibrium values and first-best values, Non-Autonomous case

Treatments	Agents	Scientist		Producer		Consumer			
		Effort	Lie (rate)	Info. (rate)	Lie (rate)	Buy if G (rate)	Buy if B (rate)	Effort if G	Effort if B
P (Total) - TNAP		3,21	0,68	0,575	0,33	0,88	0,49	3,89	4,04
P (1/2) and S (1/2) - TNAS		4,03	0,36	0,467	0,5	0,92	0,57	4,22	4,15
P (1/3) and S (1/3) and C (1/3*) - TNAC		4,05	0,52	0,483	0,48	0,6	0,38	4,31	4,54
<i>First-best</i>		2	0	0	0	1	1	+	++

A.3 Instructions

Translated from French to English.

In red the elements that change in the instructions for treatments other than T-AP.

Thank you for taking part in this experiment on decision-making. In this experiment, you have the opportunity to make money. The amount of your payoff will depend on your decisions and the decisions of other participants. Therefore, we ask you to read these instructions carefully since they will help you understand the experiment. All your decisions are anonymous. You will never type your name into the computer. You will give your choices to the computer in front of which you are sitting.

From now on, communication is no longer permitted. Please switch off your mobile phone as well. If you have a question, raise your hand and an experimenter will come and answer you in private.

This experiment comprises 3 parts. You have received the instructions for part 1. Each time you finish a part, you will get the instructions for the next one. All participants have the same instructions

The earnings you can collect by taking part in this experiment are expressed in Euros for the last two parts and in ECUS (Experimental Currency Units). At the end of each part, your earnings, in ECUs, will be converted in euros using the conversion rate $1000 \text{ ECUS} = 7.5\text{€}$.

The winnings you can accumulate by participating in this experiment are expressed directly in Euros for the last two parts and in ECUS (experimental currency units) for the first part, and will be converted into Euros using the conversion rate $1000 \text{ ECUS} = 7.5\text{€}$. At the end of the experiment, the gains you will have earned, converted into euros, will be paid to you in cash privately

PART 1

Context

This experiment involves three roles: a scientist, a producer and a consumer. The three players make their decisions sequentially. First the scientist, then the producer and finally the consumer. The scientist has the opportunity to design a technology that can be highly reliable or unreliable, depending on a certain probability. He then transfers his technology to a producer who can sell it to a consumer. The scientist and the producer will be remunerated according to the selling price of the technology. The consumer will use the technology and, depending on how it is used, will be able to make a profit from it. However, if the technology fails, it will cause damage to society, which will have to be borne equally by the scientist, the producer and the consumer.

Before you start, your role will be revealed, and each participant will retain the same role throughout the experiment. The experiment consists of 5 successive periods. During each period, you will be randomly assigned to a group of 3 players: a scientist, a producer and a consumer. You will never play two periods with the same players.

The scientist

At the beginning of each period, the scientist receives 1000 ECUS to develop a technology, which will be sold by the producer to the consumer. There are two types of technology. The technology can be highly reliable or unreliable, depending on a certain probability.

For the scientist, the probability of developing a highly reliable or unreliable technology depends on the cost of developing it. The greater the cost of developing the technology, the more likely it is to be highly reliable. The table below shows the probability of having a highly reliable technology as a function of the cost of developing it:

Investment by the Scientist	0	4	16	64	128	512	1024
Probability of G	10%	30%	40%	50%	60%	70%	80%

For example, if he devotes zero cost to developing this technology, the technology has a 10% chance of being very reliable, or a 90% chance of being unreliable. Conversely, if he spends a maximum cost of 100 ECU, the technology has a 90% chance of being highly reliable (10% chance of being unreliable).

The difference between a highly reliable and an unreliable technology is given by the probability that the technology will fail. Unreliable technology has a 60% chance of failure, and highly reliable technology has only a 20% chance of failure.

Having decided on this development cost, the scientist passes the technology on to a producer, who will then sell it to the consumer. To this end, the scientist must tell the consumer whether the technology he has developed is unreliable or highly reliable. The scientist is not obliged to tell the truth, and may, for example, state that his technology is very reliable, when in fact it is unreliable. The producer must then offer this technology to the consumer, who may or may not buy it at the price announced by the producer.

The gains for the scientist in each period will depend on the finalization of the sale of the technology from the producer to the consumer, its selling price and whether or not the technology has failed. If the producer announces a highly reliable technology, he will be able to sell it at a price equal to 1500 ECU. If he announces an unreliable technology, it will be sold at a price of 1400 ECUS. The scientist still receives 1/3 of the sale price if the consumer buys the technology from the producer.

in *T-AS* and *T-NAS*: On the other hand, if the technology fails when the consumer uses it, this results in a loss of 2400 ECU, which the scientist and the producer will have to bear in equal parts. This costs the scientist 1200 ECU.

in *T-AC* and *T-NAC*: On the other hand, if the technology fails when the consumer uses it, this results in a loss of 2400 ECU, which the scientist, the producer and the consumer will have to bear in equal parts. This costs the scientist 800 ECU.

There are two (three) possible outcomes:

- If the technology is not purchased by the consumer : Scientist's gain = 1000 - development cost
- If the technology is purchased by the consumer and does not fail: Scientist's gain = 1000 - development cost + $1/3 * \text{price}$
- in *T-AS* and *T-NAS* If the technology is purchased by the consumer and fails: Scientist's gain = 1000 - development cost - 1200 + $1/3 * \text{price}$
- in *T-AC* and *T-NAC* If the technology is purchased by the consumer and fails: Scientist's gain = 1000 - development cost - 800 + $1/3 * \text{price}$

As a reminder, the reliability of the technology depends on the cost of its development, which can vary from 0 to 100 ECU. The selling price of the technology by the producer to the consumer will be equal to 1500 ECU if the technology is advertised as very reliable, or 1400 ECU if it is advertised as unreliable.

The producer

The producer receives 1500 ECUS at the beginning of each period. He also receives the technology developed by the scientist, which he will market to the consumer. He receives information from the scientist on the reliability of the technology, but does not know whether the technology is really very reliable or unreliable. He can, however, find out the true nature of the technology by paying a cost of 50 ECU to acquire the information. He will then know whether the technology is very reliable or unreliable.

Then he sells the technology to the consumer. Whatever the reliability of the technology, he can claim to the consumer that it is a very reliable technology. So he can tell the consumer that the technology is very reliable when it's actually unreliable, and vice versa. If the technology is advertised as very reliable by the producer, he can sell it at 1500 ECUS. If the technology is advertised as unreliable, it will be sold at 1400 ECUS. If the consumer buys the technology, the scientist and the producer share the profit from the sale: $2/3$ of the price goes to the producer, and $1/3$ to the scientist.

As with the scientist, the producer's earnings in each period will depend on whether or not the producer has sold the technology to the consumer, and on the selling price of the technology. This gain will be deducted from 50 ECU if the producer has paid to know the true nature of the technology. If the technology fails, the society suffers a loss of 2,400 ECU, to be borne fully by the scientist. There are three possible outcomes:

- If the technology is not purchased by the consumer : Producer gain = 1500 - information cost
- If the technology is purchased by the consumer and does not fail: Producer gain = 1500 - information cost + $2/3 * \text{price}$
- If the technology is purchased by the consumer and fails : Producer gain = 1500 - information cost - 2400 + $2/3 * \text{price}$

- in *T-AS* and *T-NAS* If the technology is purchased by the consumer and fails : Producer gain = 1500 - information cost - $1200 + 2/3 * \text{price}$
- in *T-AC* and *T-NAC* If the technology is purchased by the consumer and fails : Producer gain = 1500 - information cost - $800 + 2/3 * \text{price}$
- in *T-AF* If the technology is purchased by the consumer and fails : Producer gain = 1500 - information cost - $2400 + 2/3 * \text{price} - 500$

As a reminder, the cost of information is 0 or 50 ECU, depending on whether or not the producer buys the information on the state of the technology. The price received is equal to 1500 ECU if the technology is advertised as highly reliable, or 1400 ECU if it is advertised as unreliable (when purchased by the consumer).

The consumer

The consumer receives 1500 ECUS at the beginning of each period. He must decide whether or not to buy the technology proposed by the producer. The technology may fail, in which case it causes a loss of 2400 ECU, the cost of which will be shared equally between scientist, producer and consumer. But in the event of failure, the technology also generates an additional loss for the consumer, as he or she derives less benefit from its use. Indeed, if the consumer buys the technology, he can use it and derives a benefit of 1800 ECUS, when no failure occurs. If, on the other hand, a failure occurs, the benefit is only 1200 ECU.

The producer tells the consumer whether the technology is highly reliable or unreliable, but the consumer has no way of knowing whether it is the type of technology advertised. As a reminder, whatever the type of technology, there is always a probability that it will fail. Simply put, a highly reliable technology is less likely to fail (20% chance) than an unreliable technology (60% chance). The probability of a technology being highly reliable, or unreliable, depends on the cost to the scientist of developing it.

In *T-NAP*, *T-AS* and *T-NAC*: That said, whatever the nature of the technology, the consumer can make an effort to reduce the likelihood of the technology failing. To do this, he must take part in a task. The task consists in counting the number of '1's in a table of '0's and '1's, as shown in the screenshot below. He has 1 minute to solve a maximum of 4 tables. For every table composed of '0' and '1' that it counts exactly, the probability of failure decreases. The following table shows the failure probabilities associated with the number of correctly counted tables, in the case of highly reliable and unreliable technology.

# of tables	0	1	2	3	4
Good technology	0.2	0.15	0.11	0.075	0.05
Bad technology	0.6	0.45	0.33	0.23	0.15

In *T-AC* and *T-NAC*: As for the producer and the scientist, if the technology fails, it causes damage to society of 2400 ECU, which the scientist, the producer and the consumer have to bear in equal parts. It then

costs the consumer 800 ECU. The consumer's earnings in each period will depend on whether or not he buys the technology and, if he does, whether or not a failure occurs.

- If the technology is not purchased by the consumer : Consumer gain = 1500
- If the technology is purchased by the consumer and does not fail: Consumer gain = 1500 - price + 1800
- If the technology is purchased by the consumer and fails: Consumer gain = 1500 - price + 1200
- in *T-AC* and *T-NAC*: If the technology is purchased by the consumer and fails: Consumer gain = 1500 - price + 1200 - 800

The purchase price is equal to 1500 ECUS if the technology is advertised as very reliable, or 1400 ECUS if it is advertised as unreliable (when purchased by the consumer).

At the end of the experiment, one period out of the 5 will be effectively remunerated according to the euro conversion rate. A participant will draw a period at random in order to calculate the payout for this first part. Each period has the same probability of being selected.

PART 2

In this part, you will have only one decision to make. You will have to choose **one** gamble from 5 different gambles. Your earnings for this part will depend on the outcome of the gamble. For each gamble, there are 2 possible earnings: earnings from situation A and earnings from situation B. Each situation has a 50% chance of happening.

In order to determine your earnings for this part, the computer will virtually toss a coin. If it is heads, situation A will happen and if it is tails, situation B will happen. Your earnings will correspond to the earnings of the winning situation of the gamble you will have chosen.

[Displayed on the screen:]

Gamble	Situation A (50%)	Situation B (50%)
1	3€	3€
2	4€	2.5€
3	5€	2€
4	6€	1.5€
5	7€	1€

PART 3

In this game, you'll be asked to make 10 successive decisions. For each of these decisions, you will be randomly associated with another participant, who will be called the passive participant, but you won't know his identity. Likewise, they won't know your identity. It's a different person for each of the 10 decisions. For

each of the 10 decisions you'll have to make, you'll have the choice between two alternatives, A and B. Each alternative has consequences for you and your partner. For example, the following table shows you a decision. You can choose between alternative A, which gives you 2 euros and the other participant 3.25 euros, or alternative B, which gives you both 2.5 euros. Which alternative do you choose?

Alternative A		Your choice		Alternative B	
You get	the other player gets	A	B	You get	the other player gets
2 €	3.5 €			2.5 €	2.5 €

You'll have 10 decisions of the same type to make. Your payout for this round will be determined at the end of the experiment. One of the ten decisions will be chosen at random, and a payment will be made according to your choice. If, for example, the decision chosen is the one given as an example in the table, and you chose alternative B, then both you and the other participant will receive 2.5 euros. As you will also be randomly associated with another participant who will have to make the same decision as you, you will also receive a payout as a passive participant following the choice made by this participant.

A.4 Final questionnaire

1. Your age
2. Your gender: 0 M, 1 F, 2 Other
3. The type of degree you are enrolled in: 0 Licence, 1 Master, 2 Doctorat
4. Your field of study: 0 Law, 1 Economics-Business, 2 Literature-Languages, 3 Exact Sciences, 4 Psycho-Socio, 5 Political Sciences, 6 Other
5. (please specify)
6. In life, do you consider yourself a risk-taker or a cautious person? or cautious? Indicate on a scale of 0 to 10 where you think you stand 0 representing a person who is extremely cautious and 10 representing a person who representing a person who loves taking risks:, 0 to 10
7. In life, would you say that most of the time, you try to help others or are you mainly concerned with your own interests? Indicate on scale from 0 to 10 where you think you stand, with 0 representing a person who love to help others and 10 representing a person acting solely in 0 to 10
8. In life, would you say that most of the time, people try to help others or only look after their own interests? Indicate on a scale of 0 to 10 where you place others, 0 representing a person who loves to help others and 10 representing a person acting solely in his or her own interest, 0 to 10
9. Generally speaking, do you feel that most people can be trusted or that we should be very careful when dealing with others? 0 Most people can be trusted, 1 We should be very careful
10. During part 3 of the experiment, what information guided your decision at each stage? 1 Your gain and that of the other so that the other has no less than you, 2 Your gain and that of the other so that the other has no more than you, 3 Your gain and that of the other so that the other has no more or less than you.
11. What criteria guided your decisions during the first part of the experiment?
12. In your opinion, what was the aim of the experiment? What do you think we wanted to test?

A.5 Calibration

In this section, we provide details about our choices as regards the calibration of the numerical values for the variables and parameters which are used in our experiment. The calibration is made on the assumption of risk-neutral agents, as the theoretical model did. All calculations are available upon request.

First, the value of harm $H = 2400$ can be divided by 2 and 3 (in case where liability is equally shared among the three agents). From this value, we set the initial endowment of each agents ($W_C = 1500, W_P = 1500, W_S = 1000$) and the consumer's benefit from using the good ($B_{sup} = 1800, B_{inf} = 1500$) to be lower than the harm. Agents' endowments are also set to be sufficiently high to ensure a strictly positive payoff, at the end of the game, in case of a harm occurring (and in case of maximum liability for that agent, i.e., $L_C = 800, W_P = 2400, W_S = 1200$ - knowing that in case of harm, the good is sold and so the Producer and the Scientist both enjoy a share of the the selling price).

As regards the probability of failure (see Table 2), we set values in a way to obtain a three factor between $p_G(\epsilon)$ and $p_B(\epsilon)$, for a given ϵ . This ensures a difference in expected cost of harm between both qualities of technology. By construction, the decrease in probability is steeper with a B -quality technology than with a G -quality one, leading the Consumer to make a higher effort, in case of a non-autonomous car, when facing a type B than when facing a type G : $\epsilon_B^* > \epsilon_G^*$. Without any cost function (since this is a cognitive task), we arbitrarily choose: $\epsilon_B = 3$ and $\epsilon_G = 1$, leading to $p_G(\epsilon_G) = 0.15$ to and $p_B(\epsilon_B) = 0.23$. (ϵ_G is set 1 below the median value, and ϵ_B is set 1 above the median value).

Our parameters respond to a first theoretical assumption: both qualities of technologies (B and G), whatever the kind of good (autonomous, or not) are socially desirable. The most stringent case is the case of an autonomous good of B -quality: our parameters ensures $B_{sup} - p_B(H + (B_{sup} - B_{inf})) \geq 0$.

As regards private decisions, our calibration responds to additional constraints and has implications. Recall that our calibrations are based on risk-neutral agents, aiming to maximize their expected profit.

First, consider the Consumer. In case of an autonomous good, our parameters values lead a risk-neutral consumer to always buy the good if he has no liability in case of harm, except when he thinks that the Producer lies (he offers a good embedded with a B -quality technology at a price ρ_G). In the case where the consumer has liability, our parameters lead the risk-neutral consumer not to buy an autonomous good embedded with a B -quality technology.

In case of a non-autonomous car (and considering $\epsilon_G = 1$ and $\epsilon_B = 3$), the risk-neutral consumer which has no liability always buys the good (whatever the quality of embedded technology), and in case of liability he buys the good except when he thinks that the Producer lies.

So, conditions seem to be more favourable for buying a non-autonomous good. However, recall that we only take into account monetary benefits and costs. So we ignore the (non-monetary cost) of making efforts for decreasing the probability of failure in case of non-autonomous good, which can also refrain from buying the good (and recall that in case of no effort, probabilities of failure are the same as in the case of an autonomous good).

Then, consider the Producer. Recall that he has to make two decisions: (i) buying or not additional information to discover the true quality of the technology (B or G), (ii) offering the good to the consumer and declaring

the quality of the technology (not observable by the Consumer).

In case of non-autonomous good, our calibration is made to ensure the (risk-neutral) Producer to invest in information seeking whatever his (strictly positive) level of liability.

In case of autonomous good, as shown by the theoretical analysis, the Producer has no interest in buying information. However, we calibrate the level of the fine $F_P = 500$ in a way to ensure the (risk-neutral) Producer to buy information. In details, we have $p_B = 0.6$ and a fine of 500. But the Producer has a belief in the probability the Scientist lying to him. Posing 0.5 for this belief, the expected fine is: $0.5 \cdot 0.6 \cdot 500 = 150$. Given a cost of 50 to acquire information, the Producer has an interest to do it.

As regards the decision about which quality of technology to announce to the Consumer, as shown by theory the Producer's decision is in line with the his about buying (or not) information: here, in case of non-autonomous good, the Producer doesn't lie (he uses the information he buys, in order the Consumer to make the appropriate level of effort), and in case of autonomous good he lies (in case of type B) except when he can be fined.

Finally, consider the Scientist. He has to make two decisions: (i) investing to increase the likelihood to design a Good technology; (ii) declaring a quality of technology to the Producer.

As regards the declaration of the quality of technology, as shown by theory, in case of an autonomous good embedded with a technology of B-quality, the Scientist always has interest in declaring a G-quality. This is the same in case of a non-autonomous good when he faces no liability. However, we calibrate our parameters to ensuring a (risk-neutral) Scientist, who faces a strictly positive liability in case of a non-autonomous good, to have interest in declaring the true quality of the technology.

Concerning the levels of investment, our parameters leads to the values mentioned in Table A.6, in Appendix A.7. We can remark that all levels of investment are lower to socially optimal ones, and that investments are lower in case of non-autonomous goods than in case of autonomous ones. This is the consequence of the lower difference in probabilities of failure in case of non-autonomous goods relative to the case of autonomous ones ($p_B(\epsilon_B) - p_G(\epsilon_G) = 0.23 - 0.15 = 0.08$ in case of non-autonomous, and $p_B - p_G = 0.6 - 0.2 = 0.4$ in case of autonomous), which leads to a lower difference in expected liability between the two types of technology. This reduces the incentives for the Scientist to design a Good technology in case of non-autonomous goods (and suppress the differences in levels of investment depending on the level of liability).

As a last remark, we also ensure that a G-quality technology (whatever fully or semi-autonomous) is always socially desirable, given the cost of investment. Here we have: $-c(e^{**}) + (p(e=0) - p(e^{**})) * (p_B - p_G) * (H + (B_{sup} - B_{inf})) > 0$, with $-c(e^{**}) = -128$, $p(e=0) - p(e^{**}) = 0.9 - 0.4 = 0.5$.

A.6 Additional predictions for risk averse and risk seeking agents

The Consumer. As for a risk neutral consumer, we can derive predictions about the probability to buy the good and the effort to make when the technology is Non-Autonomous (i.e. treatments T-NAP, T-NAS and T-NAC) for agents having other risk preferences. If we assume risk-averse agents, we can easily deduce that their net utility from using the car is higher in case of a G-quality technology than in case of a B-quality technology. This is so because the expected benefit from a Good technology is higher than the expected benefit from a Bad one. If the technology is non autonomous and the consumer is not liable at all, the price of a good embedded with a G-technology is 100 ECUS higher than the price of good with a B-technology. But a G-quality allows to save 120 in terms of expected harm and is associated with a lower level of risk (variance) than a B-quality. As a result, a risk-averse Consumer has a higher likelihood to buy the good when facing a G-quality than when facing a B-quality.

However, compared to a risk-neutral agent, it is not sure that a risk-averse Consumer has an interest in buying the good, since consuming the good is a source of risk while not buying it leads the Consumer to keep her endowment with certainty: there is a trade-off between risk and expected revenue.

Also, if we depart from risk neutrality and assume a risk-averse Consumer, the possibility to reduce the level of risk of accident, even at a given (cognitive) cost, may provide a higher utility from buying the good when the technology is Non-Autonomous rather than Autonomous. This means that for a given level of liability, a risk-averse consumer may be more prone to buy a good endowed with a Non-Autonomous technology than a good endowed with an Autonomous technology.

The Producer. We have shown above that in case of a non-Autonomous technology, the risk-neutral Producer invests in information seeking and the benefit from paying for information increases with the share of liability. Also, in this case, the Producer declares the true quality of technology to the Consumer. On the contrary, if the technology is autonomous, the Producer does not pay for information and always announces a G-quality technology, except when a fine applies in case of failure for declaring a B-quality technology (following an accident). In that case, the Producer pays for information and announces the true quality.

If instead we assume a risk-averse Producer, the decision to pay for information and to announce the true quality to the Consumer will depend on what the producer expects to be the Consumer's behavior. Knowing that a sufficiently risk-averse Consumer may decide not to buy a good endowed with a B-quality technology, even a risk-averse Producer may decide not to buy information and announce a G-quality technology to maximize the likelihood of selling the good. But if the Producer thinks the Consumer to be few risk-averse, who would be prone to buy a good endowed with a B-quality technology then, in case of Non-Autonomous good, the Producer will pay and will announce the true quality. In that case acquiring information is beneficial

and leads to a reduction in the level of risk (variance).³⁴

The Scientist. The predictions are also slightly affected by risk preferences. In particular, we can discuss what should be the behaviors of the Scientist in the presence of risk-aversion for all agents. As regards which quality the Scientist declares to the Producer, the same rationale applies that for the declaration of the Producer to the Consumer. The decision of the Scientist to declare the true quality (or to lie) depends on what behavior to expect from other agents, and in particular from the Consumer. If, by telling the truth, the Scientist thinks that it is (too) likely that the risk-averse Consumer chooses not to buy the good (and assuming that the Producer also chooses to tell the truth to the Consumer), then the risk-averse Scientist may prefer to lie. But if the Scientist expects the Consumer to be few risk-averse, in a way to be prone to buy a good endowed with a B-quality technology and to make effort to reduce the probability of causing harm (Non-Autonomous good), then the risk-averse Scientist has incentives to declare the true type, to push the Consumer to reduce the level of the risk.

Concerning the decision to invest to increase the chance of developing a G-quality technology, we know that a risk-neutral Scientist makes decisions in a way to maximize the expected payoff. As a result, increasing the level e of investment beyond the level chosen by the risk-neutral Scientist leads to a decrease in expected payoffs for the risk-averse Scientist. However, knowing that a G-quality technology allows to reduce the probability of accident, in the case where the Scientist bears some degree of liability a risk-averse Scientist may invest more than a risk-neutral one to the extent that a G-quality technology allows to reduce the level of risk (variance).

A.7 Summary of predictions

The Table A.6 summarizes, for each treatment (i.e., each combination of sharing of liability and type of technology (Autonomous, or Non-Autonomous)), the theoretical predictions about private decisions made by each agent, given our specifications and parameters values. We also add, in the last column, the first-best decisions which are associated with this framework.

“A” refers to the Autonomous technology, and “NA” to the Non-Autonomous one. “Lie” means that the agent declares the technology to be of G-quality when he knows the technology to be of B-quality and/or when he is not sure about what is the true quality (Producer). “Truth” means the agent to declares the true quality, eventually after having searched information about it (Producer).

³⁴The minimum net gain in expectation from investing in information is met in the case where $l_P = 800$. It is : $0.22*800 - 50 - 2/3*100 = 59.33$, i.e., the decrease in expected damages (by using a G-quality instead of a B-quality), minus the cost of acquiring information and the cost of a lower selling price. It reaches 411.33 when $l_P = 2400$. 0.22 is the difference between $p_B(\epsilon_G^*) = 0.45$ and $p_B(\epsilon_B^*) = 0.23$.

Table A.6: Predictions and equilibria

Agent	Technology	Decision	$l_p = 1$	$l_p = l_s = 1/2$	$l_p = l_s = l_c = 1/3$	$l_p = 1+\text{fine}$	First-best
Scientist	A	e Announcement	$e = 2$ Lie	$e = 3$ Lie	$e = 2$ Lie	$e = 2$ Lie	$e = 4$ Truth
	NA	e Announcement	$e = 1$ Lie	$e = 1$ Truth	$e = 1$ Truth	- -	$e = 2$ Truth
Producer	A	Information Announcement	No Lie	No Lie	No Lie	Yes Truth	- Truth
	NA	Information Announcement	Yes Truth	Yes Truth	Yes Truth	- -	- Truth
Consumer	A	Buy	Yes	Yes	Yes	Yes	Yes
	NA	Buy Effort if G-quality Effort if B-quality	Yes Low Medium	Yes Low Medium	Yes Medium High	- - -	Yes High Very high