# Opening up the Black Box: Technological Transparency and Prevention

Lu Li [*]

December 6, 2019

## Abstract

We discuss the behavioral and welfare implications of uncovering the causal mechanism of prevention. We introduce the concept of technological transparency (TT) – the extent to which scientific knowledge reveals the mechanism of prevention. While TT improves welfare through more efficient preventive efforts, this improvement may be undermined or reversed if information is incompletely disclosed or if the risk is insurable. TT affects behavior through an ex-ante information channel and an ex-post regret channel. Our findings inform the cost-benefit analysis of advancing the knowledge about risk determinants, the effective disclosure of such knowledge, and the design of information campaigns to promote public safety.

**Keywords:** technological transparency · prevention · value of information · regret

**JEL-Classification:** D61 · D80 · D90

[*] Ludwig-Maximilians-University Munich, Institute for Risk Management and Insurance, E-Mail: li@bwl.lmu.de, Phone: +49 089 2180 3929.

Electronic copy available at: https://ssrn.com/abstract=3501523

## 1 Introduction

*"Success = talent + luck. Great Success = a little more talent + a lot of luck"*

- Daniel Kahneman, *Thinking, Fast and Slow*

Most life outcomes depend inevitably on both our own actions and factors beyond our control. Disentangling the roles of luck and effort, however, is often not trivial. In any situation where the efficacy of effort corresponds to a change in the probability of some particular outcome, the exogenous determinants of the outcome are completely hidden. A prominent example of such is self-protection (also referred to as loss prevention, see Ehrlich and Becker, 1972; Courbage et al., 2013), which is a costly effort to reduce the likelihood of a loss event.

While healthy diet and regular physical exercise help reduce the probability of developing diabetes, the success of healthy lifestyle in preventing diabetes is shown to depend on various factors including one's genetic makeup (Frayling, 2007). However, the mechanism of this gene-lifestyle interaction is still far from being perfectly understood (Qi et al., 2008).

As in the example mentioned above, any self-protection technology has an inherent possibility of failing. While an agent knows by how much a larger effort is less likely to fail, she does not know what makes her effort succeed or fail. In other words, the underlying mechanism of self-protection resembles a black box in its conventional interpretation. How can we open up this black box and what happens if we open it up?

In this article, we propose the concept of *technological transparency* (TT), which refers to the extent to which scientific knowledge allows people to understand the causal mechanism of a technology. In self-protection, full TT refers to understanding all causal determinants of a risk. We model full TT by introducing a new, causal interpretation of self-protection that reveals the latent states of the decision problem from its conventional, reduced-form definition. Mapping the actual intensity of a hurricane to the minimum amount of reinforcement to keep a house safe is an example of full TT when the intensity of the hurricane is the only exogenous risk determinant.

An improvement of TT, on the other hand, corresponds to revealing previously unknown risk determinants together with how they collectively predict the likelihood for the effort to succeed. In disease prevention, the efficacy of preventive healthcare often depends on an individual's genetic makeup, which, for diabetes prevention, is shown to be associated with multiple genes scattered across an individual's genome (Ali, 2013). In this case, TT is gradually improved through the discovery of relevant genes together with their interaction with the preventive effort until the mechanism of prevention is completely understood.

After establishing the concept of TT, we study the behavioral and welfare implications of TT on self-protection according to whether the risk determinants are ex-ante or ex-post observable. When combined with ex-ante observable risk determinants, that is, the agent learns the realization of the risk determinants before making her decision, an improvement of TT refines the information partition and leads to a Pareto improvement of welfare through enabling better-informed decision making. However, this welfare improvement may be undermined or even reversed if information regarding the newly uncovered risk determinants is incompletely disclosed, that is, the agent learns her risk type but not how her risk type affects the marginal productivity of her effort. Such incomplete disclosure of information is not uncommon in reality: in the preventive healthcare for complex diseases where risks depend on both genetic and lifestyle factors, studies on the gene-lifestyle interaction are sparse compared to those studying either genes or lifestyle in isolation as risk factors (Qi et al., 2008). As personal genetic testing services become increasingly affordable and popular, people often pay to have their genetic risk factors tested. However, such tests seldom offer any information on how the revealed genetic risk types interact with the effectiveness of prevention effort.

The welfare effect of TT becomes very different when the agent can not only invest to prevent the risk, but also insure it via the (private) insurance market. With insurance, even full TT has an ambiguous effect on welfare. This is because as TT unravels the risk, it also unravels the insurance market for that risk and hence introduces a negative distributional consequence on social welfare. As a result, the overall welfare effect of TT involves a tradeoff between more efficient prevention and an exaggeration of social disparity.

We then discuss the situation where risk determinants are ex-ante unobservable. For instance, the intensity of earthquakes are extremely difficult to predict. When combined with risk determinants that are only observable ex-post, TT obviously no longer creates informational benefit for the decision making. However, even in this case, TT still reveals the optimal effort conditional on the (unobservable) value of the risk determinant. Hence, as soon as the agent observes the risk determinants ex post, she realizes what she "should have done" in hindsight, although at this point it would be already too late to change her effort. We argue that this hindsight effect induced by TT can alter the agent's choice of effort by

Electronic copy available at: https://ssrn.com/abstract=3501523

triggering counterfactual thinking and regret, that is, disutility from realizing having made a suboptimal decision (Loomes and Sugden, 1982; Bell, 1982). We show that anticipating future regret raises the self-protection effort and that the more regret-averse the agent is, the stronger the effect of TT. This result also shows that even if TT is currently not yet available, there exists a positive effect on effort from anticipating future acquisition of TT.

Our findings suggest that more efficient and adequate risk mitigation effort may be enabled by advancing the knowledge about previously hidden risk determinants. However, in order to fully exploit the value of such knowledge and in extreme cases, prevent such knowledge from harming welfare, it is crucial for the interaction between the preventive effort and the newly discovered risk determinants to be revealed to the decision-maker in addition to the risk determinants themselves. In case of insurable risks, the positive welfare effect of TT is no longer guaranteed unless an additional wealth redistribution is imposed. Our results shed light on the effective communication of scientific discoveries about risk determinants to the public. They also inform the evaluation of how much benefit results from understanding the mechanism of preventive activities.

The rest of this paper is structured as follows. Section 2 defines technological transparency (TT) based on a new way of modeling self-protection. Section 3 analyzes the consequences of TT when risk determinants are ex-ante observable. Section 4 examines ex-post observable risk determinants and the indirect behavioral impact of TT through anticipated future regret. Section 5 concludes and discusses the implications of our results. All proofs appear in the appendix.

# 2 Self-Protection and Technological Transparency

## 2.1 Self-Protection

Consider an agent with an initial endowment $w > 0$ who faces a potential loss $L$, $L \in (0, w)$ being a positive constant. First, let us recap the conventional definition of self-protection that is commonly adopted by prior literature.

**Definition 1.** (Self-protection: a reduced-form model)
Self-protection is a costly effort that reduces the loss probability. Let $x$ denote the cost of self-protection. The loss probability $p(x)$ is decreasing in $x$.

An inherent property of self-protection is that any effort may either succeed or fail[1]. While prior self-protection analyses assume $p(x)$ to be decreasing and convex in $x$ with almost no exception, they are silent regarding when and why the effort will succeed or fail. We aim to discuss exactly the determinants of success while preserving the properties of $p(x)$ commonly adopted by prior studies. To do so, we introduce a new definition of self-protection by revealing the latent states of the decision problem from its conventional interpretation.

**Definition 2.** (Self-protection: a causal model)
Consider a probability space $(\Omega, \mathcal{F}, \mu)$ where $\Omega$ is the state space, $\mathcal{F}$ is the $\sigma$-algebra and $\mu$ is the probability measure. Let $x$ denote the cost of self-protection. The loss $l(\omega, x)$ is a random variable with the support $\{0, L\}$. $l(\omega, x)$ is non-increasing in $x$ for all $\omega \in \Omega$.

**Definition 3.** A reduced-form self-protection model is said to *represent* a causal self-protection model if $p(x) = \mu\left(\{\omega \in \Omega \mid l(\omega, x) = L\}\right)$ for all $x$.

Definition 2 revisits self-protection through a causal, mechanismic lens. In each state of the world, no risk should exist and the effort has to be the sole determinant of the occurrence of the loss. In addition, the loss size is non-increasing in effort, that is, as long as some effort suffices (fails) to prevent the loss, then in the same state, any higher (lower) effort will also suffice (fail) to prevent the loss. When interpreting self-protection as a reduction of the loss probability as in Definition 1, the states are hidden and the state space is merely partitioned into two events: "loss" and "no loss". Since the states within either event are indistinguishable from one another, the mechanism of self-protection is hidden in its conventional interpretation. Definition 2, on the other hand, distinguishes between the states and allows us to disentangle the roles of nature and effort in determining the occurrence of the loss.

Definition 3 connects every causal model with a representing reduced-form model[2]. Specifically, the loss probability in the representing reduced-form model coincides with the collective probability of all states in the causal model where the loss occurs despite the effort – i.e., states that are "bad enough" for the effort not to succeed. In fact, the causal model im-

[1] We assume $p(x)$ to be strictly between 0 and 1. However, all conclusions in this paper remain unaffected if $p(x) = 1$ or $p(x) = 0$ are allowed for some effort levels.

[2] Every causal model is represented by a unique reduced-form model, but the reverse is generally not true. To see this, imagine swapping two states $\omega_1$ and $\omega_2$ with identical probability, which will change the causal model but not the reduced-form model.

Electronic copy available at: https://ssrn.com/abstract=3501523

plies the existence of a random variable that characterizes the states exactly in terms of their desirability.

**Corollary 1.** *For every causal self-protection model, there exists a threshold effort: a random variable $t\colon \Omega \to \mathbb{R}$ satisfying $l(\omega, x) = L \cdot \mathbb{I}\{x < t(\omega)\}$, where $\mathbb{I}\{\cdot\}$ is the indicator function. Furthermore, the threshold effort follows a mixed type distribution whose survival function coincides with $p(\cdot)$ in the reduced form model that represents the causal model. Equivalently, the cumulative distribution function $F(\cdot)$ of the threshold effort satisfies:*

- $F(t) = 0, \qquad \text{if } t < 0$
- $F(t) = 1 - p(t), \qquad \text{if } t \geq 0.$

The threshold effort is the lowest effort such that the loss does not occur. It is a manifestation of the state variable $\omega$ as a behaviorally relevant concept: the better the state, the lower the threshold effort. More effort reduces the probability of the loss event by exceeding more potential realizations of the threshold effort and, as a result, pushing more states into the no loss event.

The states, while described as a purely abstract concept in the causal model, are in fact determined by exogenous risk determinants such as the weather, an individual's genes, the type of virus causing the next influenza, or the combination of multiple risk determinants in case more than one exist. In reality however, it is common for the $p(x)$ function to be obtained, e.g. through randomized controlled experiments, before the actual risk determinants are uncovered. As scientific research gradually uncovers the hidden risk determinants behind $p(x)$, we move closer towards understanding the exact casual mechanism of prevention.

Now suppose there are altogether $N$ risk determinants[3] $\tilde{y}_1, \tilde{y}_2, \cdots, \tilde{y}_N$ whose ranges are denoted by $Y_1, Y_2, \cdots, Y_N$, respectively. We have the following definition:

**Definition 4.** In a causal self-protection model,

- *full technological transparency* means knowing the invertible function $\lambda : \prod_{i=1}^{n} Y_i \to \Omega$;
- *an improvement of technological transparency* induced by $\tilde{y}_i$ means knowing the function $\lambda_i : Y_i \to \mathcal{F}$ such that $\lambda_i(y) = \left\{\omega \in \Omega \mid \lambda^{-1}(\omega) \cdot \hat{e}_i = y\right\}$, where $\hat{e}_i$ stands for the column unit vector whose $i$th row equals 1.

[3] We use a tilde sign to indicate a random variable.

Technological transparency (TT henceforth) refers to the ability to identify the true state through risk determinants. The acquisition of full TT requires the identification of all risk determinants together with the estimation of $l(\omega, x)$ for each state[4]. The following example (simplified for an expositional purpose only) demonstrates the concepts of TT, the state, and the threshold effort.

**Example 1.** An agent wants to reinforce her house to prevent it from being destroyed by a hurricane. By conducting experiments and simulations, researchers show that the success of the reinforcement effort depends solely on the intensity of the hurricane and that each potential intensity requires a minimum effort such that the house is kept safe, i.e. the threshold effort.

In Example 1, TT is obtained by identifying the (only) risk determinant and mapping it to the state, which is represented by the corresponding threshold effort. Hence, a higher reinforcement effort is more likely to succeed by exceeding the threshold efforts of more potential intensities of the hurricane.

In case the risk has more than one determinant, it is common for the risk determinants to be gradually uncovered one after another. The discovery of additional, but not all risk determinants leads to an improvement of TT, which we illustrate using the following example.

**Example 2.** The development of disease A depends on both the quality of one's lifestyle $x$ and one's genetic makeup. People start by realizing the benefit of $x$, but not knowing with which exact gene(s) $x$ jointly determines disease A. One day, scientific research shows that gene $\tilde{y}_1$ is relevant for disease A. However, taken together, $x$ and $\tilde{y}_1$ still do not completely explain the occurrence of A. 5 years later, research further uncovers gene $\tilde{y}_2$ as another risk determinant. When taking $x$, $\tilde{y}_1$ and $\tilde{y}_2$ into account, the occurrence of disease A can now be completely predicted.

For simplicity, let us assume that both genes have 2 potential variants: $Y_1 = \{L, H\}$ and $Y_2 = \{l, h\}$. Then, full TT would reveal that there are altogether four possible states in this problem, each being mapped from a particular combination of genes: $\omega_1 = \lambda\left((L, l)\right)$, $\omega_2 = \lambda\left((L, h)\right)$, $\omega_3 = \lambda\left((H, l)\right)$, and $\omega_4 = \lambda\left((H, h)\right)$. However, in Example 2, TT is gradually obtained in two steps. When only $\tilde{y}_1$ is uncovered, we obtain the function $\lambda_1$ such that

[4] TT corresponds to discovery rather than foreknowledge according to Hirshleifer (1971)'s seminal work on the social value of information. Foreknowledge refers to the ability to predict things that are currently perceived as stochastic but will eventually reveal themselves to all observers, such as tomorrow's weather. Discovery, on the other hand, refers to uncovering hidden properties of nature, such as physical laws.

Electronic copy available at: https://ssrn.com/abstract=3501523

$\lambda_1(L) = \{\omega_1, \omega_2\}$ and $\lambda_1(H) = \{\omega_3, \omega_4\}$. $\lambda_1(L)$ and $\lambda_1(H)$ represent two different "risk types", but for each risk type, disease A is still jointly determined by $x$ and "something else" that is unknown at this stage. It is only when $\tilde{y}_2$ is also uncovered that we have a complete understanding of the mechanism of $x$.

## 3 Ex-ante Observable Risk Determinants

In this section, we analyze the behavioral and welfare consequences of TT assuming all risk determinants are observable prior to the time the decision is made. We address ex-ante unobservable risk determinants in Section 4[5].

First, let us review the concept of information partition introduced by Aumann (1976)[6]. let $\mathcal{P}$ denote an information partition: a partition of $\Omega$ into subsets containing subjectively indistinguishable elements. Mathematically, $\mathcal{P}$: $\Omega \to \mathcal{F}$ is a function that maps every $\omega \in \Omega$ into $\mathcal{P}(\omega) \subset \Omega$. If $\omega$ is the true state, then the agent regards all states within $\mathcal{P}(\omega)$ as possible and all states outside $\mathcal{P}(\omega)$ as impossible. The finer the information partition, the more capable the agent is of distinguishing between states and the closer she is to knowing the true state.

**Corollary 2.** *When risk determinants are ex-ante observable,*

- *full TT implies $\mathcal{P}(\omega) = \{\omega\}$ for all $\omega \in \Omega$;*
- *an improvement of TT induced by $\tilde{y}_i$ implies $\mathcal{P}(\omega) = \{\hat{\omega} \in \Omega \mid \lambda_i^{-1}(\hat{\omega}) \cdot \hat{e}_i = \lambda_i^{-1}(\omega) \cdot \hat{e}_i\}$, where $\hat{e}_i$ denotes the column unit vector whose $i$th row equals 1.*

Together with ex-ante observable risk determinants, full TT corresponds to the finest information partition where every element is a singleton, in other words, the ability to identify the exact true state of the world. In the context of Example 2, when both genes $\tilde{y}_1$ and $\tilde{y}_2$ are uncovered and agents can conduct genetic tests to learn their genotype, the state space is partitioned into $\{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}\}$. An agent whose test result is $L$ and $l$ then identifies herself in state $\omega_1 = \lambda(\{L, l\})$. Another agent who is tested as the combination of $L$ and $h$, for instance, then identifies $\omega_2 = \lambda(\{L, h\})$ instead.

[5] As the development in big data and predictive analytics makes forecasting services increasingly available and accurate, the results in this section also reflects the value of forecasting (see Blair and Romano, 1988; Katz and Murphy, 2005, for instance).

[6] Information partition has also been applied to model financial literacy, see Neumuller and Rothschild (2017).

Electronic copy available at: https://ssrn.com/abstract=3501523

An improvement of TT, on the other hand, corresponds to further refining the information partition along the value of the newly uncovered risk determinant: states in the same partition element are those that share the same value of that risk determinant. in Example 2, when only $\tilde{y}_1$ is shown to be a risk determinant and agents can take genetic tests to learn whether they are of risk type $L$ or $H$, the state space is partitioned into $\{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\}$ and risk type $L$ $(H)$ corresponds to the first (second) partition element. Hence, this refined information partition leads to an update of the states' probabilities.

Now bring effort $x$ back to the picture. The improvement of TT results in – via updating the states' probabilities – also an update of the loss probability:

**Definition 5.** Let $\mathcal{P}$ be the information partition after an improvement of TT induced by $\tilde{y}$ and let $\omega$ be the true state. Given effort $x$, the *posterior loss probability* $p(x, y)$ is the loss probability conditional on the partition element $\mathcal{P}(\omega)$:

$$p(x, y) = \frac{\mu(\{\omega' \in \mathcal{P}(\omega) | l(\omega', x) = L\})}{\mu(\mathcal{P}(\omega))}. \tag{1}$$

Before the improvement of TT, the loss probability $p(x)$ may be seen as the average loss probability across all risk types in the population[7] given effort $x$. The improvement of TT classifies the population into different risk types along the value of the risk determinant that induces it. For each risk type, the same prevention activity is represented by a different posterior loss probability.

## 3.1 Full Technological Transparency

Let us now examine the behavioral and welfare implications of full TT given ex-ante observable risk determinants.

In the benchmark scenario, we have the classic self-protection problem where the agent optimizes her effort knowing nothing but $p(x)$. We assume $x$ to be the amount of disutility induced by the self-protection effort. We also assume $p(x)$ to be twice differentiable and satisfy $p' < 0$ and $p'' > 0$. The choice of the optimal self-protection effort corresponds to the

[7] We assume for now $p(\cdot)$ to be unbiased and discuss potential biasedness of the prior estimation of $p(\cdot)$ in Section **??**.

Electronic copy available at: https://ssrn.com/abstract=3501523

following optimization problem:

$$\max_x U(x) = [1 - p(x)]u(w) + p(x)u(w - L) - x, \tag{2}$$

the solution of which, denoted by $x^0$, is determined by the first order condition:

$$-p'(x^0)[u(w) - u(w - L)] = 1. \tag{3}$$

The left hand side of (3) represents the expected marginal benefit per unit of effort and the right hand side represents the marginal cost. Define $\hat{x} = u(w) - u(w - L)$, which is the utility premium induced by the loss (Friedman and Savage, 1948). It is easy to see that $\hat{x}$ corresponds to the highest effort the agent is willing to undertake to reduce the loss probability from 1 to 0. Hence, (3) can also be written as:

$$p'(x^0) = -\frac{1}{\hat{x}}, \tag{4}$$

The second order condition is fulfilled given the model assumptions.

Now assume there is full TT. How will this affect the agent's optimal effort?

**Proposition 1.** *Under full TT, when the risk determinants are observed ex-ante, the optimal self-protection effort equals the threshold effort if the threshold effort does not exceed $\hat{x}$; it equals zero otherwise. Moreover, full TT increases effort with probability $p(x^0) - p(\hat{x})$ and decreases effort with probability $1 - p(x^0) + p(\hat{x})$.*

The intuition underlying Proposition 1 is straightforward: Full TT reveals the true threshold effort before the agent makes her choice. Hence, whenever the loss is preventable, she chooses to prevent the loss at the lowest possible cost[8]. If the loss is not preventable, she chooses not to exert any effort. Note that we use the term unpreventable to refer to two situations: either the threshold effort is infinite or it is finite but too costly, that is, it costs more than the gain from avoiding the loss, which is the utility premium. Since both situations have the same behavioral consequence, we do not distinguish between them for the sake of convenience. Depending on the revealed threshold effort, full TT may either increase or decrease effort. In fact, full TT and the risk determinants collectively correspond to a

[8] We restrict ourselves to problems where the loss either does or does not occur exactly once.

Electronic copy available at: https://ssrn.com/abstract=3501523

conclusive information structure. Such an information structure fully removes risk from the decision problem by providing perfect information, which induces the most efficient decision (see Gollier, 2001; Blackwell, 1951).

We now turn to welfare. Consider $p(x)$ to be the unbiased estimation of loss probability given effort $x$ in the entire population. After normalizing the number of individuals in the population to 1, we have:

**Proposition 2.** *Full TT:*

- *leads to a Pareto improvement in welfare.*
- *increases utilitarain social welfare by* $\hat{x}p(x^0) + x^0 - \int_0^{\hat{x}} p(t)\mathrm{d}t > 0$*, out of which*
  - *(a)* $\hat{x}p(x^0) - x^0 p(\hat{x}) - \int_{x^0}^{\hat{x}} p(t)\mathrm{d}t$ *results from eliminating underprevention;*
  - *(b)* $x^0 - \int_0^{x^0} p(t)\mathrm{d}t$ *results from reducing the cost of prevention;*
  - *(c)* $p(\hat{x})x^0$ *from saving costs from losses that are not preventable.*

Obviously, (a), (b) and (c) are mutually exclusive and collectively comprehensive subsets of the population. Since individuals belonging to either category undergo a welfare improvement, full TT implies a Pareto improvement in welfare. While the number of deaths due to insufficient prevention is often used to emphasize the importance of promoting preventive healthcare (see for instance Danaei et al., 2009; Keeney, 2008), (b) suggests that this number may be reduced by TT [9]. However, in addition to leaving no more preventable losses unprevented, TT also generates value through saving costs that would be otherwise wasted independent of the loss being preventable or not.

The improvement of social welfare also has a straightforward graphical representation. It is illustrated by the overall area of the shaded regions in Figure 1, where the blue curve represents $p(x)$, the dashed green line is the tangent line at $x^0$ that is parallel to the solid green line connecting the points $(0, 1)$ and $(\hat{x}, 0)$, indicating the fulfillment of (4). In particular, $A$ stands for the welfare gain from saving the cost of prevention, $B_1$ and $B_2$ jointly represent the benefit from eliminating underprevention, and $C$ is the benefit from not wasting effort on unpreventable losses.

The following corollary identifies a condition related to technological improvement (see Hoy and Polborn, 2015; Lee, 2015) that leads to a higher welfare improvement by full TT.

[9] see also Baillon et al. (2019) for a discussion of underprevention with probability weighting

Electronic copy available at: https://ssrn.com/abstract=3501523

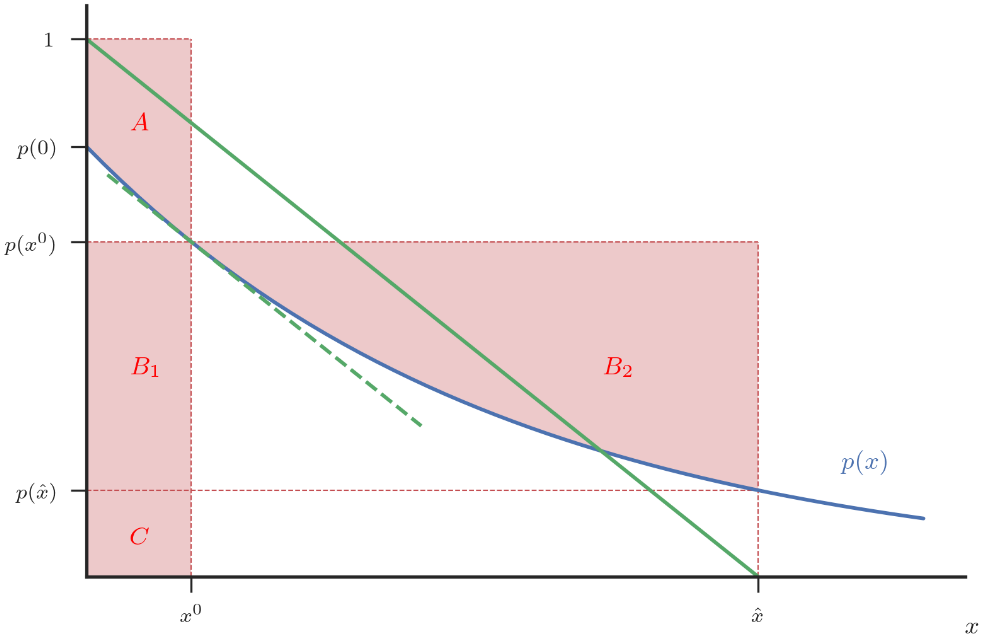

Figure 1: Full TT improves utilitarian social welfare. $x^0$ is the optimal effort in the benchmark case. $\hat{x} = u(w) - u(w-L)$. The total area of the shaded regions equals the improvement of social welfare by full TT. $A$ stands for the welfare gain from reducing the cost of successful prevention. $B_1$ and $B_2$ jointly represent the benefit from eliminating underprevention. $C$ is the benefit from not wasting effort on unpreventable losses.

**Corollary 3.** *The welfare improvement by full TT increases when the self-protection technology improves from* $p(x)$ *to* $q(x)$ *such that* $q(x) < p(x), \forall x \neq x^0$, $q(x^0) = p(x^0)$ *and* $q'(x^0) = p'(x^0)$.

As illustrated by Figure 2, whenever there is a global technological improvement such that the marginal productivity of effort is preserved at $x^0$ and the loss probability becomes lower for all efforts except $x^0$, one can expect a larger welfare improvement by full TT. It is easy to see that such a technological improvement is equivalent to a deterioration of the threshold effort in the sense of first order stochastic dominance (FSD). Interestingly, without TT, the technological improvement has neither behavioral nor welfare consequence. This is because the optimal effort in the benchmark case is determined solely by the local condition (4), which completely ignores the improved outcome for every effort other than $x^0$. However, full TT takes advantage of the deterioration of every potential realization of the threshold effort and enhances the value of the technological improvement by allowing the latter to be fully (globally) exploited.

Electronic copy available at: https://ssrn.com/abstract=3501523

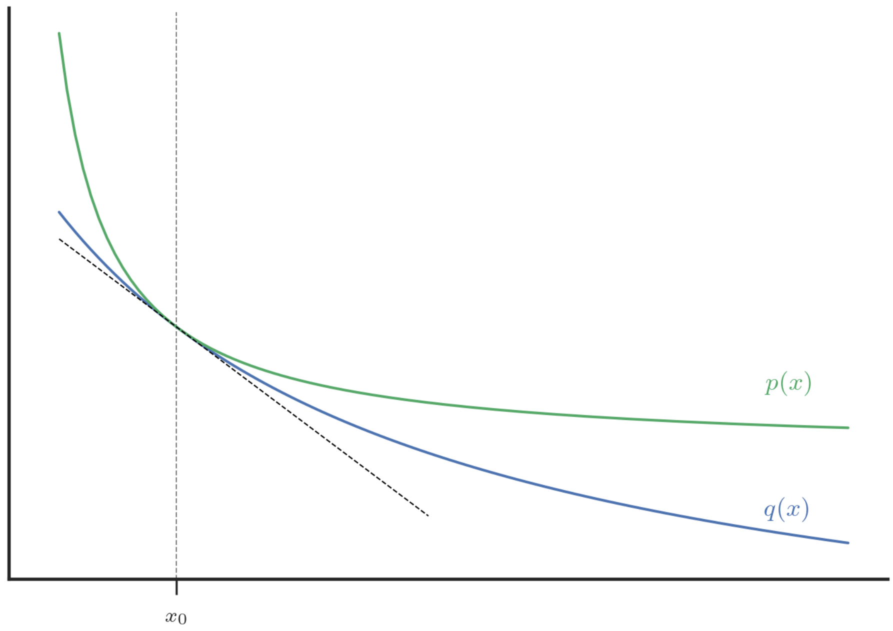

Figure 2: A technological improvement that increases the welfare improvement of full TT.

## 3.2 An Improvement of TT

In this section, we demonstrate that while the results in 3.1 qualitatively carry over to partial improvements of TT, they may be undermined or even reversed by the incomplete disclosure of information, which is in fact not uncommon in reality. We also relax the assumption of an unbiased prior $p(x)$ and examine the impact of biased prior beliefs on the value of improving TT.

Many risks are a result of highly complex interactions of a number of risk determinants. This is particularly true for complex diseases involving the interaction of genetics and lifestyle factors (see Willett, 2002, for instance). As a result of the complex nature of risks, researchers are often still in the process of gradually improving TT. We study the generic case where an improvement of TT is induced by uncovering one additional risk determinant, but the generality of our model allows the results to be easily extendable to an improvement of TT induced by uncovering multiple risk determinants.

Consider an improvement of TT induced by the discovery of the risk determinant $\tilde{y}$. This results in the posterior loss probability $p(x, y)$. Assume $p_1(x, y) < 0$, $p_{11}(x, y) > 0$, $y \in [\underline{y}, \overline{y}]$, and $\delta(x) = p(x, y) - p(x, y') > 0$ whenever $y' > y$ so that a larger $y$ corresponds to a smaller

Electronic copy available at: https://ssrn.com/abstract=3501523

revealed risk level. Following the terminology of Hoy (1989), we call $\delta(x)$ the difference function with respect to the risk determinant $\tilde{y}$. Although we do not impose monotonicity on $\delta(x)$, in cases where $\delta'(x)$ does have a consistent sign, we may obtain some interesting unambiguous comparative statics[10]. We follow Hoy (1989) and make the following distinction:

**Definition 6.** The self-protection technology has:

- increasing difference (ID) with respect to $\tilde{y}$ if $\delta'(x) > 0$;
- constant difference (CD) with respect to $\tilde{y}$ if $\delta'(x) = 0$;
- decreasing difference (DD) with respect to $\tilde{y}$ if $\delta'(x) < 0$;

Definition 6 characterizes the relationship between the productivity of the effort and the revealed risk type. In particular, ID (DD) means the effort is more effective when the risk is high (low), whereas CD means the effectiveness is independent of the risk type. Note that in contrast to Hoy (1989) and Li and Peter (2019), the property of our difference function is specific to the revealed risk determinant $\tilde{y}$: if a technology has DD with respect to $\tilde{y}$, it may still have ID, CD or a non-monotonic difference function with respect to other risk determinants. Figure 3 illustrates properties of $\delta(x)$.

When there is no knowledge about risk determinants, $p(x)$ is commonly obtained through either randomized experiments or simply through experience. Experience, however, may be strongly subject to bias. The next definition addresses the potentially biased estimation of the prior loss probability $p(x)$. Suppose $\tilde{y}$ is distributed in the population with the cumulative distribution function $G(y)$. Then,

**Definition 7.** $p(x)$ is said to be unbiased if $p(x) = \int_{\underline{y}}^{\overline{y}} p(x, y) \mathrm{d}G(y)$.

In other words, for each $x$, the unbiased loss probability must correspond to the population average of the posterior loss probabilities. A common reason for its violation is the so-called selection bias (Harrison and List, 2004; Cleave et al., 2013; Allcott, 2015), that is, the sample used in the estimation of $p(x)$, either via randomized controlled trials (RCTs) or through the experience of some individuals, is not representative of the entire population. Selection

[10] In fact, there are cases where $\delta(x)$ cannot be monotonic. An extreme example is full TT, where the conditional loss probability reduces to 1-0 step functions as mentioned in the previous subsection. Under full TT, $\delta(x)$ itself is also a step function that starts at 0 from the left of the $x$ axis, then jumps to 1 at the lower threshold effort, and eventually jumps back to 0 at the larger threshold effort.

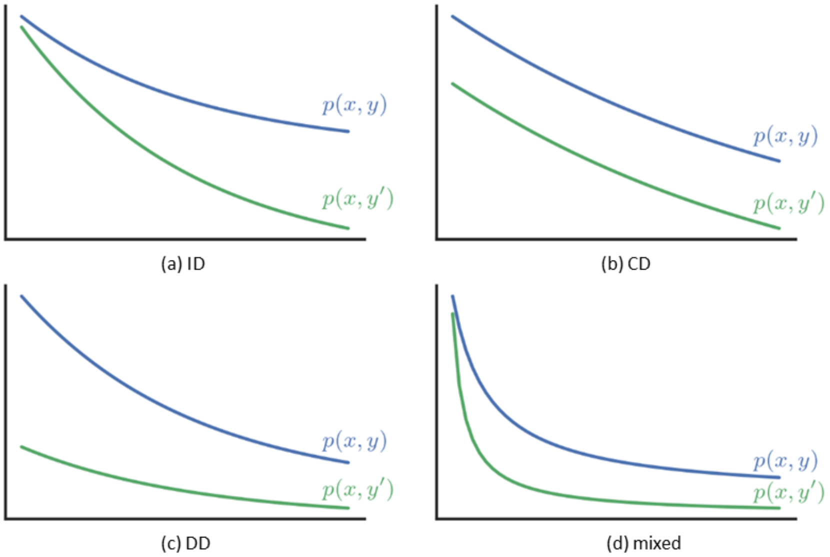

Figure 3: Properties of the difference function $\delta(x) = p(x, y) - p(x, y')$ where $\underline{y} \leq y < y' \leq \overline{y}$. (a), (b) and (c) represent ID, CD and DD, respectively. (d) stands for a mixed case with ID for small $x$'s and DD for large $x$'s.

bias may occur, for instance, because participants of RCTs can only be recruited via certain channels such as universities or local clinics, because some groups of participants are more likely than others to self-select into the studies, or because of the deliberate exclusion of certain population groups (such as women in their pregnancy) due to ethical or liability concerns (Shields and Lyerly, 2013).

**Proposition 3.** *Consider an improvement of TT induced by the risk determinant $\tilde{y}$. Let $x(y)$ denote the optimal effort conditional on $\tilde{y} = y$. It holds that:*

- *$x(y)$ increases (decreases) with $y$ if the technology has ID (DD) with respect to $\tilde{y}$.*
- *$x(y) \equiv x^0$ if the technology has CD with respect to $\tilde{y}$.*

*Furthermore, the improvement of TT:*

*(a) leads to a Pareto improvement in welfare as well as improved utilitarian social welfare if the technology has ID or DD with respect to $\tilde{y}$, but improves welfare no more than full TT does.*

*(b) does not affect welfare if the technology has CD with respect to $\tilde{y}$;*

*(c) may reduce welfare if properties of the difference function are misperceived.*

Electronic copy available at: https://ssrn.com/abstract=3501523

(d) *leads to higher welfare improvement when the estimation of $p(x)$ is subject to selection bias than when it is not.*

The first part of Proposition 3 focuses on the behavioral impact of an improvement of TT and is consistent with Hoy's result with 2 risk types. As a direct consequence of (4), it depends on the interaction between the revealed risk level and the effectiveness of prevention. The optimal effort always balances the marginal cost and the marginal benefit. Since the former is constant, the relationship between the optimal effort and the revealed risk level is entirely determined by how the marginal productivity changes with $y$, which is governed by exactly the difference function. Under DD, a worse revealed risk, or a lower $y$, means higher marginal benefit, which leads to an increase of effort, whereas the opposite is true under ID. In disease prevention where $\tilde{y}$ is an individual's genetic makeup, evidence shows that a higher effort often attenuates the influence of genes, suggesting DD is more likely to apply (see Qi et al., 2008, 2012; Graff et al., 2017, for instance). In this case, individuals with high-risk genes are expected to exert higher effort after learning their risk profile through a genetic test.

The second part of Proposition 3 addresses the welfare impact of an improvement of TT. An improvement of TT is equivalent to an imperfect signal that is less informative than full TT, but still serves to improve the quality of decision-making and therefore increase social welfare, although to a lesser extent than full TT does. In particular, welfare only gets improved if observing the signal changes one's optimal action, which is true for ID and DD, but not for CD.

An important implication of Proposition 3 is that the knowledge about the revealed risk determinant, such as the result of a genetic test, will not necessarily improve the quality of one's decision unless the test result is communicated along with knowledge about the gene-effort interaction. Many personal genetic testing services charge extra money for revealing individuals' health risk profile, e.g. by reporting whether the tested individual possesses genetic variants that increase the risk of type 2 diabetes. However, these test results say nothing about how the revealed genetic variants affect the marginal productivity of preventive effort. Suppose DD is true, meaning that high risks should exert more effort after obtaining their test result. However, if people do not also know that DD applies to the revealed genes, they might mistakenly perceive ID to be the case and reduce their effort instead if they see themselves as being "too unfortunate to benefit from anything". The latter is particularly

Electronic copy available at: https://ssrn.com/abstract=3501523

relevant if the cost of effort is high, such as when fast food becomes increasingly cheap to produce and easy to access, or when one's work environment imposes a sedentary lifestyle.

As the incomplete disclosure of the improved knowledge creates leeway for biased beliefs to form, people may feel they are learning something about themselves, but end up making worse choices than before learning the information: a harmful "illusion of knowledge". Hollands et al. (2016) show that revealing DNA based risk estimates through personal genetic tests does not lead to significant behavioral change. We argue that this none-finding may be partly explained by incomplete disclosure of information. As argued by Qi et al. (2008), medical research on the gene-lifestyle interaction in disease prevention is scarce compared to those targeting genes or lifestyle in isolation. Proposition 3 shows that studying the interaction as well as communicating the findings of these studies to the public has crucial importance in turning research findings into real value.

Moreover, the more biased the prior knowledge $p(x)$, the more value the improvement of TT generates. The selection bias leads to a suboptimal choice of effort in the benchmark case due to the fact that $p(x)$, which equals $\int_{\underline{y}}^{\overline{y}} p(x,y)\mathrm{d}H(y)$, yields a biased estimation of the marginal benefit. If the sample is biased towards better risks, the estimated marginal benefit is also closer to that of the better risks. Under DD, this means a smaller marginal benefit and hence a benchmark effort that is too low. However, since the selection bias does not affect welfare after the improvement of TT, the improvement of TT must lead to higher welfare improvement than without the selection bias. Proposition 3 suggests that whenever ill-informed decision making exists due to selection bias, the suboptimal choice can be corrected by an improvement of TT. This is particularly meaningful for situations where the cost for improving TT is lower than the cost of eliminating selection bias.

### 3.3 Insurable Risks

We now extend the welfare analysis to situations where the risk can be insured through private insurance markets under symmetric information. We focus on full TT, but derive results that qualitatively also apply for partial improvements of TT. Hoy (1989) analyzed the welfare effect of increasing the precision of risk categorization for two given risk types under different insurance contracting possibilities. In a way, the analysis in this section also extends Hoy (1989) to the most extreme case where any risk type is fully decomposed into a combination of infinitely many degenerate risks.

Consider first the benchmark case. The agent faces the joint decision of both her preventive effort $x$ and the level of insurance. We assume a perfectly competitive market where the insurer earns zero profit and the premium is actuarially fair. Under symmetric information, the insurer may observe the level of prevention the agent undertakes and prices it into the insurance contract. Formally, let $\alpha$ stand for the level of coinsurance, that is, the agent receives $\alpha L$ from the insurer in case the loss occurs. The premium she pays upfront equals the expected indemnity payment $p(x)\alpha L$, which decreases with her preventive effort. First of all, it is easy to see that the agent always opts for full insurance ($\alpha = 1$) conditional on any effort level since the elimination of the risk is available at fair cost. Her decision therefore reduces to a uni-variate optimization problem:

$$\max_x U(x) = u(w - p(x)L) - x$$

As before, we let $x^0$ denote the optimal effort in the benchmark case, which satisfies the first-order condition

$$u'(w^0)p'(x^0)L = -1$$

where $w^0 = w - p(x^0)L$.

Let us now switch to full TT, where the loss is solely determined by the agent's effort. In this case, the risk is no longer existent, which is why the insurance market for the risk also no longer exists [11]. Therefore, the level of insurance is always zero ($\alpha = 0$) and the optimal effort is exactly as described in Proposition 1. If an agent finds out her threshold effort equals zero, she enjoys a loss-free world without having to exert any effort, which is obviously better than the benchmark case where she has lower wealth due to paying the insurance premium $p(x^0)L$ and a non-zero expenditure on prevention $x^0$. The case is similar if the threshold effort is positive but comparably low. At the other end of the spectrum, if the loss is unpreventable, then under full TT, the agent faces a sure loss without being able to do anything about it, whereas in the benchmark case, the loss still occurs but gets fully indemnified by her insurance contract, which she purchased at a lower price than her actual (degenerate) risk deserves. Obviously, for such extremely "unlucky" agents, full TT reduces welfare. Therefore, full TT can no longer lead to a Pareto welfare improvement. From the utilitarian perspective, TT

[11] The fact that the insurance market no longer exists under full TT also holds under asymmetric information.

Electronic copy available at: https://ssrn.com/abstract=3501523

introduces two effects. First, it introduces a distributional effect by unraveling the insurance market, which favors the "lucky" but harms the "unlucky" individuals, therefore exaggerating social disparity and harming social welfare. The distributional effect of TT also resonates with Hirshleifer (1971)'s well-known result on the social value of information when agents can trade state-contingent claims. At the same time, TT also makes prevention more efficient as discussed in Section **??**, which in itself increases social welfare. As a result, the overall net welfare effect of full TT depends on the tradeoff between the efficiency effect and the distributional effect.

**Proposition 4.** *When the risk is insurable, full TT*

- *never leads to a Pareto improvement of welfare;*
- *may increase or decrease utilitarian social welfare. The net welfare effect of full TT equals* $u(w) - u(w^0) + x^0 - \int_0^{\hat{x}} p(t)\mathrm{d}t$.

The impact of full TT on utilitarian social welfare is illustrated by the difference between $A$ and $B$ in Figure 4, where $\check{x} = u(w) - u(w^0) + x^0$ represents the welfare gain resulting from full TT if the threshold effort equals 0, or the largest threshold effort for full TT to increase welfare. Obviously, the larger $\check{x}$ gets, the larger (smaller) $A$ ($B$) becomes and the more likely full TT will lead to an increase of social welfare.

## 4 Ex-Ante Unobservable Risk Determinants and Regret

We now turn to situations where the risk determinants are ex-ante unobservable. Such situations are very common. For instance, one only knows the exact intensity of the next earthquake after the earthquake occurs. In the penalty kick of a soccer game, the goalkeeper only realizes his opponent's strategy at the end of the kick.

When risk determinants are only observable ex-post, TT no longer affects a rational, forward-looking agent's choice since it neither removes nor signals the risk ex-ante. However, we argue that by revealing the threshold effort conditional on the realization of the ex post observable risk determinants, full TT may trigger ex post regret as the agent observes the outcome and realizes what she "should have done" in the past. Regret, if anticipated and incorporated into the decision-making ex-ante, serves as a second, indirect channel for TT to affect behavior.

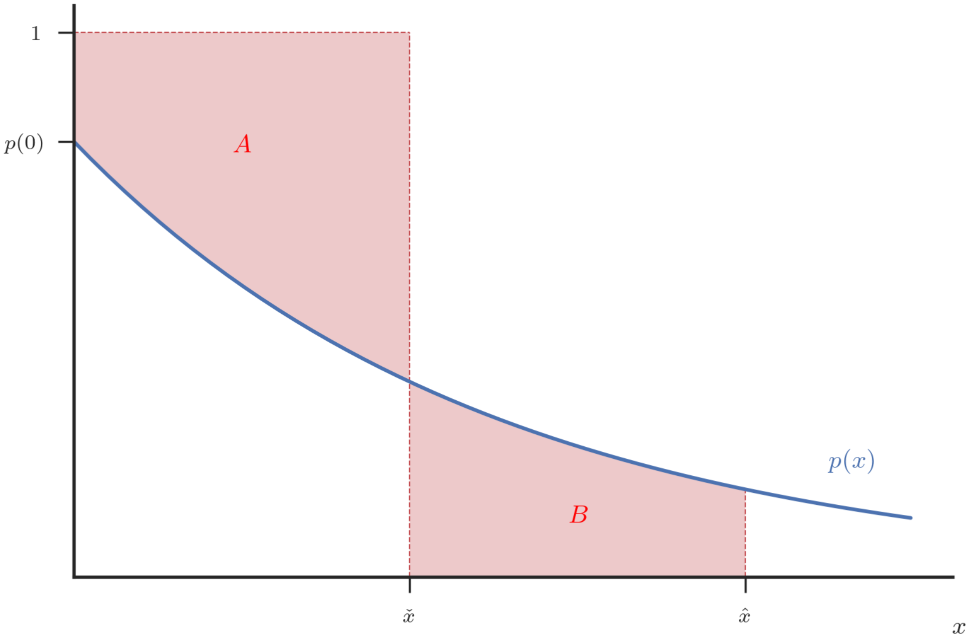

Figure 4: The impact of full TT on utilitarian social welfare equals $A - B$ when insurance is available. $\hat{x} = u(w) - u(w - L)$. $\check{x} = u(w) - u(w - p(x^0)L) + x^0$.

Ample evidence shows people are not always forward looking as assumed by classic decision theories. Regret is a concept well documented by psychologists since more than a century ago (see Zeelenberg and Pieters, 2007, for a survey). Regret theory, which is initially discussed in the economic literature by Bell (1982); Loomes and Sugden (1982); Fishburn (1982), assumes people experience disutility from realizing having made a suboptimal choice (see Bleichrodt et al., 2010; Camille et al., 2004, for existing empirical support for regret theory). While the original regret theories are restricted to decision problems with 2 alternative actions, Sugden (1993) and Quiggin (1994) generalize the theory to arbitrary choice sets based on a set of axioms. More recently, regret theory has been applied to various economic decisions including insurance demand (Braun and Muermann, 2004), actions (Engelbrecht-Wiggans, 1989; Engelbrecht-Wiggans and Katok, 2008) and portfolio choice (Muermann et al., 2006; Muermann and Volkman Wise, 2006) and is shown to explain observed deviations from predictions of the expected utility theory including the Allais paradox, preference for low deductible insurance contracts and the disposition effect.

Since the self protection problem has a continuous choice set, we follow Braun and Muermann (2004) and adopt the approach of regret theoretical expected utility (RTEU) as the

Electronic copy available at: https://ssrn.com/abstract=3501523

basis of our analysis. The RTEU approach features an arbitrary choice set and is consistent with both Sugden (1993)'s axiomatic approach and Quiggin (1994)'s Irrelevance of Statewise Dominated Alternatives (ISDA) assumption. It assumes regret is expressed as a function of the difference between the utility that would be obtained from the foregone optimal decision and the utility obtained from the actual decision:

$$\psi(x,s) = \phi(x,s) - k \cdot g\left[\phi\left(x^{opt}\left(s\right), s\right) - \phi(x,s)\right], \tag{5}$$

where $x$ is the choice variable, $s$ is the realization of the random state variable, $\phi$ is the classic von Neumann-Morgenstern utility as a function of the choice and the state (also referred to as choiceless utility), $x^{opt}(s)$ is the foregone optimal action given the realized state $s$, $g$ with $g > 0$,$g' > 0$, $g'' > 0$ and $g(0) = 0$ represents regret, and $k \geq 0$ stands for the intensity of regret aversion. Hence, the decision problem is written as:

$$\max_{x} \mathbb{E}\psi(x,s) = \mathbb{E}\left\{\phi(x,s) - k \cdot g\left[\phi\left(x^{opt}\left(s\right), s\right) - \phi(x,s)\right]\right\}, \tag{6}$$

Notably, without TT and the causal self-protection model, the concept of regret seems almost incompatible with the self-protection problem: the agent would never be able to find out the foregone optimal decision since the states within the loss (no loss) event are not distinguishable from each other. However, $x^{opt}$ becomes revealed if full TT is available in combination with the ex post observation of the risk determinants. Consider again Example 1 regarding the hurricane. Suppose there is full TT, that is, the decision-maker knows the threshold effort given each potential intensity of the hurricane as well as the probability distribution of the intensity. Based on this information, she chooses effort $x^0$ that protects her house from a hurricane up to the intensity $y_1^0 = t^{-1}(x^0)$. In addition, she anticipates four potential future scenarios: (1) The hurricane is weaker than $y_1^0$, her house remains safe, but she would have obtained the exact same outcome had she put in less effort. (2) The hurricane has exactly the intensity $y_1^0$ and her house remains safe. (3) The hurricane is stronger than $y_1^0$ and destroys her house, but she would have avoided the loss had she chosen a higher effort. (4) The hurricane is so strong that no effort would be able to prevent it. Any effort undertaken would be therefore completely wasted. Whichever scenario occurs, the agent will realize in hindsight what she should have done, which, except for in the second scenario, differs from what she actually did and therefore generates regret. If the ex post regret is anticipated ex-ante, it will in turn

Electronic copy available at: https://ssrn.com/abstract=3501523

affect the optimal effort since the agent wants to additionally mitigate the expected amount of regret. Formally, the scenarios above are summarized by the following objective function:

$$\begin{aligned}\max_x V(x) =&F(0)\left[u(w)-x-k\cdot g(x-0)\right]+\int_0^x \left[u(w)-x-k\cdot g(x-t)\right]f(t)\mathrm{d}t\\ &+\int_x^{\hat{x}}\left[u(w-L)-x-k\cdot g(u(w)-u(w-L)+x-t)\right]f(t)\mathrm{d}t\\ &+\int_{\hat{x}}^{\infty}\left[u(w-L)-x-k\cdot g(x-0)\right]f(t)\mathrm{d}t, \end{aligned}\tag{7}$$

The first line of (7) is the RTEU of situations where the actual effort exceeds the threshold effort, no loss occurs and the decision-maker regrets spending too much effort. Note that the first line is decomposed into two parts due to the discontinuous distribution of the threshold effort at 0. The second line stands for when the actual effort is lower than the threshold effort, the loss occurs and the agent regrets spending too little effort. The third line is when the loss is unpreventable and the decision-maker regrets spending any effort at all. Note that if we eliminate $g$ from (7), it collapses to the original decision problem in our benchmark case described by (2). The optimal effort of a regret-averse decision-maker $x^r$ is therefore determined by the first order condition of (7), which is obtained by applying the Leibniz rule:

$$\begin{aligned}V'(x^r) =& f(x^r)\hat{x}+f(x^r)kg(\hat{x})-\int_0^{x^r} kg'(x^r-t)f(t)\mathrm{d}t-\int_{x^r}^{\hat{x}} kg'(\hat{x}+x^r-t)f(t)\mathrm{d}t\\ &-\left[1-F(x^r)\right]kg'(x^r)-F(0)kg'(x^r)-1\\ =&\ 0.\end{aligned}\tag{8}$$

By evaluating the sign of $V'(x^0)$, we can compare the optimal effort of a regret-averse decision-maker with that of an expected utility maximizer.

**Proposition 5.** *With full TT and ex post observable risk determinants, the demand for self-protection increases with regret aversion.*

Generally speaking, an increase of effort is always associated with two types of marginal benefits and two types of marginal costs. On the one hand, since the loss probability becomes further reduced, the decision-maker is both more likely to obtain higher wealth and less likely to regret letting a preventable loss occur. On the other hand, the effort itself costs more and the amount of regret increases due to the higher sunk cost. Taken together, when evaluated at $x^0$, the net effect of higher effort is positive because the strongest regret comes from realizing

Electronic copy available at: https://ssrn.com/abstract=3501523

having spent too little effort, and the convexity of $g$ makes the agent disproportionally averse to large regrets. Hence, anticipating future regret makes her willing to undertake more self-protection ex-ante. TT plays an essential role in this process by revealing the threshold effort, which is the crucial reference point without which a regret-averse agent would not be able to objectively attribute the observed event to internal (her effort) or external (the realizations of the risk determinants) causes.

An interesting related question is how a regret-averse agent would behave when TT is unavailable or when some risk determinants are never observable[12]. One possible answer to this question is regret is irrelevant when the foregone optimal decision is unknown, as assumed in Engelbrecht-Wiggans and Katok (2008). However, counterfactual thinking has long been documented in the psychology literature (see Zeelenberg, 1999, for a survey). Phenomena such as the hindsight bias (Christensen-Szalanski and Willham, 1991), outcome bias (Baron and Hershey, 1988), or different attributional styles (Abramson et al., 1978) also suggest there is an innate tendency for people to subjectively assign reasons to past events even when they do not possess adequate information to do so. We therefore believe regret may nevertheless have an impact on behavior without TT, although the direction of this impact may be largely subject to contextual and behavioral factors. For instance, the amount of regret may be much higher for an extremely pessimistic agent who always attributes failures to herself and successes to luck than an extremely optimistic agent who believes in the exact opposite. Incorporating biased beliefs into the analysis also requires understanding whether a decision-maker subject to biased beliefs can foresee the bias ex-ante (see the similar distinction between naïve and sophisticated present bias and self control in O'Donoghue and Rabin, 1999; Ali, 2011). We leave these questions to future research but believe our analysis serves as a benchmark against which the consequences of biased beliefs may be evaluated.

## 5 Conclusion and Outlook

In this article, we propose the concept of technological transparency (TT) by explicitly interpreting the mechanism of self-protection as the interactive determination of the occurrence of

[12] Bell (1983) analyzed regret with unknown consequence of the foregone action when the choice set is binary, where he discusses a potential willingness to pay to avoid resolving the outcome of the alternative action. More recently, Gabillon (2018) extends the discussion to an arbitrary choice set. Both discussions are based on pre-assumed normative conditions on the decision-maker's preference.

Electronic copy available at: https://ssrn.com/abstract=3501523

a loss by the preventive effort and other risk determinants. Based on a novel reinterpretation of self-protection from a causal, mechanismic lens, TT translates the probability of the loss event into the joint distribution of risk determinants, which manifests itself in the distribution of the state variable and the threshold effort.

When risk determinants are ex-ante observable, full TT corresponds to the acquisition of perfect information and induces the most efficient prevention. Under full TT, every preventable loss is prevented at the lowest cost and no cost is wasted on unpreventable losses. In addition, full TT may enhance the value of technological improvements by allowing the latter to be fully (globally) exploited. A marginal improvement of TT also increases the efficiency of prevention, although to a lesser extent than full TT does. However, the positive welfare effect of an improvement of TT is no longer guaranteed if the interaction between the revealed risk type and the effectiveness of prevention is not disclosed, as the interaction crucially determines the posterior optimal effort. Improving TT also has the potential of neutralizing biased estimations of the prior loss probability. When the risk is insurable, the welfare effect of TT is ambiguous as TT disrupts the insurance market and introduces a distributional effect that exaggerates social disparity.

When the risk determinants are observed ex post, TT allows the agent to objectively attribute the (non-)occurrence of a loss to herself and to external causes. This hindsight makes the ex-ante self-protection effort increase with the degree of regret aversion. In addition, TT in combination with ex-post observable risk determinants also serves to prevent potential biased beliefs from distorting the impact of regret aversion.

Our findings suggest that more efficient and adequate risk mitigation may be enabled by scientific research uncovering risk determinants and their interaction with preventive effort. In addition, the value of such scientific research may be enhanced by technological progress reducing the cost of measuring the risk determinants, as well as predictive analytics helping forecast unknown values of the risk determinants. Our framework allows a straightforward assessment of the welfare benefit resulting from understanding causal determinants of risks, which informs cost-benefit analyses upon making resource allocation decisions, especially when the cost of improving TT is high.

Our results also have implications for the design of public education campaigns aiming to promote preventive activities. To make prevention more efficient, not only should the mechanism of a prevention technology be made transparent by research, it needs to be made

Electronic copy available at: https://ssrn.com/abstract=3501523

transparent in the eyes of the decision-maker. Instead of communicating the effectiveness of prevention based on population average statistics, policymakers may consider tailoring the information to different subpopulations to the extent allowed by current knowledge so that each subpopulation may choose the effort most suitable to their needs. In particular, for knowledge to turn into real value, it is crucial to disclose not only the risk level of each subpopulation, but also the efficacy of prevention associated with the risk level. Surprisingly, the regret channel shows that even when individuals are not yet perfectly aware of which subpopulation they belong to due to the non-observability of risk determinants, simply anticipating knowing this in the future suffices to increase the current preventive effort.

In case of insurable risks with ex-ante observable risk determinants, our results indicate that advancing our understanding of the mechanism of prevention – while making prevention more efficient – may eventually make it impossible for private insurance to exist. For such risks, an additional wealth redistribution may be necessary to recover the welfare loss resulting from the distributional effect introduced by TT.

While our discussion centers around prevention, the concept of TT has in fact a much broader range of application. Just as in self-protection, any effort that affects the outcome by altering the probability of *some* event(s) is subject to TT, where the model framework and results we developed may also be applied.

In our analysis for ex-ante unobservable risk determinants and regret, we argue that TT may trigger regret by revealing the counterfactual outcome, which is the essential reference point in regret theory. Likewise, counterfactual outcome is also a key concept of salience theory (Bordalo et al., 2012), which argues that the utility of an outcome is evaluated according to its contrast to the counterfactual outcomes instead of in isolation[13]. Future research could further explore the possible impact of TT in the context of salience theory.

In addition, while this paper focuses on the benefit side of TT, the optimal investment in improving TT has to consider its cost as well. A natural extension would be to treat the improvement of TT as an endogenous decision by incorporating the cost of acquiring TT, (see the literature on rational inattention, e.g. Sims, 2006, for the cost of information acquisition.). Another extension would be to relax the assumption that TT is pure public information and allow it to be privately obtained.

[13] Due to this common feature, salience theory and regret theory have also been compared against each other by Herweg and Müller (2019).

Electronic copy available at: https://ssrn.com/abstract=3501523

## References

Abramson, L. Y., Seligman, M. E., and Teasdale, J. D. (1978). Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology*, 87(1):49.

Ali, O. (2013). Genetics of type 2 diabetes. *World Journal of Diabetes*, 4(4):114.

Ali, S. N. (2011). Learning self-control. *The Quarterly Journal of Economics*, 126(2):857–893.

Allcott, H. (2015). Site selection bias in program evaluation. *The Quarterly Journal of Economics*, 130(3):1117–1165.

Aumann, R. J. (1976). Agreeing to disagree. *The Annals of Statistics*, pages 1236–1239.

Baillon, A., Bleichrodt, H., Emirmahmutoglu, A., Jaspersen, J. G., and Peter, R. (2019). When risk perception gets in the way: Probability weighting and underprevention. *Operations Research*, forthcoming.

Baron, J. and Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, 54(4):569.

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations Research*, 30(5):961–981.

Bell, D. E. (1983). Risk premiums for decision regret. *Management Science*, 29(10):1156–1166.

Blackwell, D. (1951). Comparison of experiments. Technical report, HOWARD UNIVERSITY Washington United States.

Blair, R. D. and Romano, R. E. (1988). The influence of attitudes toward risk on the value of forecasting. *The Quarterly Journal of Economics*, 103(2):387–396.

Bleichrodt, H., Cillo, A., and Diecidue, E. (2010). A quantitative measurement of regret theory. *Management Science*, 56(1):161–175.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly journal of economics*, 127(3):1243–1285.

Braun, M. and Muermann, A. (2004). The impact of regret on the demand for insurance. *Journal of Risk and Insurance*, 71(4):737–767.

Camille, N., Coricelli, G., Sallet, J., Pradat-Diehl, P., Duhamel, J.-R., and Sirigu, A. (2004). The involvement of the orbitofrontal cortex in the experience of regret. *Science*, 304(5674):1167–1170.

Christensen-Szalanski, J. J. and Willham, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 48(1):147–168.

Cleave, B. L., Nikiforakis, N., and Slonim, R. (2013). Is there selection bias in laboratory experiments? the case of social and risk preferences. *Experimental Economics*, 16(3):372–382.

Courbage, C., Rey, B., and Treich, N. (2013). Prevention and precaution. In Dionne, G., editor, *Handbook of Insurance*, chapter 8, pages 185–204. Springer.

Danaei, G., Ding, E. L., Mozaffarian, D., Taylor, B., Rehm, J., Murray, C. J., and Ezzati, M. (2009). The preventable causes of death in the united states: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. *PLoS Medicine*, 6(4):e1000058.

Ehrlich, I. and Becker, G. (1972). Market insurance, self-insurance, and self-protection. *Journal of Political Economy*, 80(4):623–648.

Engelbrecht-Wiggans, R. (1989). The effect of regret on optimal bidding in auctions. *Management Science*, 35(6):685–692.

Engelbrecht-Wiggans, R. and Katok, E. (2008). Regret and feedback information in first-price sealed-bid auctions. *Management Science*, 54(4):808–819.

Fishburn, P. C. (1982). Nontransitive measurable utility. *Journal of Mathematical Psychology*, 26(1):31–67.

Frayling, T. M. (2007). Genome–wide association studies provide new insights into type 2 diabetes aetiology. *Nature Reviews Genetics*, 8(9): 657–662.

Friedman, M. and Savage, L. (1948). The utility analysis of choices involving risk. *Journal of Political Economy*, 56(4):279–304.

Gabillon, E. (2018). When choosing is painful: A psychological opportunity cost model. *Working Paper*.

Gollier, C. (2001). *The Economics of Risk and Time*. The MIT Press, Cambridge, MA.

Graff, M., Scott, R. A., Justice, A. E., Young, K. L., Feitosa, M. F., Barata, L., Winkler, T. W., Chu, A. Y., Mahajan, A., Hadley, D., et al. (2017). Genome-wide physical activity interactions in adiposity—a meta-analysis of 200,452 adults. *PLoS Genetics*, 13(4):e1006528.

Harrison, G. W. and List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 42(4):1009–1055.

Herweg, F. and Müller, D. (2019). Regret theory and salience theory: Total strangers, distant relatives or close cousins? *CESifo Working Paper*.

Hirshleifer, J. (1971). The private and social value of information and the reward to inventive activity. *American Economic Review*, 61(4):561–574.

Hollands, G. J., French, D. P., Griffin, S. J., Prevost, A. T., Sutton, S., King, S., and Marteau, T. M. (2016). The impact of communicating genetic risks of disease on risk-reducing health behaviour: systematic review with meta-analysis. *The BMJ*, 352:i1102.

Hoy, M. (1989). The value of screening mechanisms under alternative insurance possibilities. *Journal of Public Economics*, 39(2):177–206.

Hoy, M. and Polborn, M. K. (2015). The value of technology improvements in games with externalities: A fresh look at offsetting behavior. *Journal of Public Economics*, 131:12–20.

Katz, R. W. and Murphy, A. H. (2005). *Economic value of weather and climate forecasts*. Cambridge University Press.

Keeney, R. L. (2008). Personal decisions are the leading cause of death. *Operations Research*, 56(6):1335–1347.

Electronic copy available at: https://ssrn.com/abstract=3501523

Lee, C.-M. (2015). Technological advances in self-insurance and self-protection. *Economics Bulletin*, 35(3):1488–1500.

Li, L. and Peter, R. (2019). Should we do more when we know less?: Optimal risk mitigation under technological uncertainty. *Working Paper*.

Loomes, G. and Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, 92(368):805–824.

Muermann, A., Mitchell, O. S., and Volkman, J. M. (2006). Regret, portfolio choice, and guarantees in defined contribution schemes. *Insurance: Mathematics and Economics*, 39(2):219–229.

Muermann, A. and Volkman Wise, J. (2006). Regret, pride, and the disposition effect. *Working Paper*.

Neumuller, S. and Rothschild, C. (2017). Financial sophistication and portfolio choice over the life cycle. *Review of Economic Dynamics*, 26:243–262.

O'Donoghue, T. and Rabin, M. (1999). Doing it now or later. *American Economic Review*, 89(1):103–124.

Qi, L., Hu, F. B., and Hu, G. (2008). Genes, environment, and interactions in prevention of type 2 diabetes: a focus on physical activity and lifestyle changes. *Current Molecular Medicine*, 8(6): 519–532.

Qi, Q., Chu, A. Y., Kang, J. H., Jensen, M. K., Curhan, G. C., Pasquale, L. R., Ridker, P. M., Hunter, D. J., Willett, W. C., Rimm, E. B., et al. (2012). Sugar-sweetened beverages and genetic risk of obesity. *New England Journal of Medicine*, 367(15):1387–1396.

Quiggin, J. (1994). Regret theory with general choice sets. *Journal of Risk and Uncertainty*, 8(2):153–165.

Shields, K. E. and Lyerly, A. D. (2013). Exclusion of pregnant women from industry-sponsored clinical trials. *Obstetrics & Gynecology*, 122(5):1077–1081.

Sims, C. A. (2006). Rational inattention: Beyond the linear-quadratic case. *American Economic Review*, 96(2):158–163.

Sugden, R. (1993). An axiomatic foundation for regret theory. *Journal of Economic Theory*, 60(1):159–180.

Willett, W. C. (2002). Balancing life-style and genomics research for disease prevention. *Science*, 296(5568):695–698.

Zeelenberg, M. (1999). Anticipated regret, expected feedback and behavioral decision making. *Journal of Behavioral Decision Making*, 12(2):93–106.

Zeelenberg, M. and Pieters, R. (2007). A theory of regret regulation 1.0. *Journal of Consumer Psychology*, 17(1):3–18.

Electronic copy available at: https://ssrn.com/abstract=3501523

# A Appendix: mathematical proofs

## A.1 Proof of Corollary 1

*Proof.* The existence of the threshold effort follows immediately from the assumption that $l(\omega, x)$ is non-increasing in its second argument. Since $\tilde{t} \geq 0$, $F(t) = 0$ whenever $t < 0$. If $t > 0$, $F(t) = Prob(\tilde{t} \leq t) = 1 - p(t)$ by definition. Furthermore, since $F(0) = 1 - p(0) > 0$, $F$ is continuous and twice differentiable everywhere except at 0. Hence, the threshold effort follows a mixed type distribution that is discrete at 0 and continuous everywhere else. It follows that the probability density function $f(t) = F'(t) = -p'(t)$ whenever $t \neq 0$. At 0, $F$ is discontinuous and $f$ does not exist. □

## A.2 Proof of Corollary 2

Corollary 2 follows naturally from Definition 4 and the definition of information partition. The proof is therefore omitted.

## A.3 Proof of Proposition 1

*Proof.* The decision problem given the threshold effort $t$ corresponds to the following:

$$\max_x U(x;t) = u(w)(1 - \mathbb{I}\{x < t\}) + u(w - L)\mathbb{I}\{x < t\} - x,$$

The solution is $t$ whenever $t < \hat{x} = u(w) - u(w - L)$ and 0 otherwise. To obtain the probabilities of TT increasing or decreasing effort from the benchmark, consider the following three scenarios:

Scenario 1. When $t \leq x^0$, the loss is prevented at the cost $t$ under full TT, whereas in the benchmark case it is prevented at the cost $x^0$. According to Proposition **??**, this happens with the probability $F(x^0) = 1 - p(x^0)$.

Scenario 2. When $x^0 < t \leq \hat{x}$, the loss is prevented at the cost $t$ under full TT, whereas in the benchmark case it occurs despite the cost $x^0$. This happens with the probability $F(\hat{x}) - F(x^0) = p(x^0) - p(\hat{x})$.

Scenario 3. When $t > \hat{x}$, the loss is unpreventable and no effort is spent under full TT, whereas in the benchmark case the cost $x^0$ is wasted. This happens with the probability $1 - F(\hat{x}) = p(\hat{x})$.

Full TT decreases effort in the first and the third scenario. It increases effort in the second scenario. □

## A.4 Proof of Proposition 2

*Proof.* For those in the population whose threshold effort is lower than $x^0$, who correspond to $1 - p(x^0)$ of the population, the aggregate utility in the benchmark case equals $W_1^{bc} = (1 - p(x^0))(u(w) - t)$, whereas the aggregate welfare under full TT equals $W_1^{TT} = u(w) -$

Electronic copy available at: https://ssrn.com/abstract=3501523

$\int_0^{x^0} tf(t)dt$. These are individuals who successfully prevent the loss in both cases, but at a lower cost under full TT. For them, the aggregate welfare improvement equals the difference between $W_1^{TT}$ and $W_1^{bc}$, which, after applying Corollary 1 and integration by parts, equals

$$\Delta W_1 = W_1^{TT} - W_1^{bc} = x^0 - \int_0^{x^0} p(t)\mathrm{d}t. \tag{9}$$

Those in the population whose threshold effort lies between $x^0$ and $\hat{x}$ suffer from underprevention in the benchmark case, so $W_2^{bc} = (p(x^0) - p(\hat{x}))(u(w-L) - x^0)$. Under full TT, they prevent the loss at the cost of $t$, so the aggregate welfare equals $W_2^{TT} = u(w) - \int_{x^0}^{\hat{x}} t\mathrm{d}t$. Hence, we have

$$\Delta W_2 = W_2^{TT} - W_2^{bc} = \hat{x}p(x^0) - x^0 p(\hat{x}) - \int_{x^0}^{\hat{x}} p(t)\mathrm{d}t. \tag{10}$$

Finally, for those with a threshold effort above $\hat{x}$, the loss is unpreventable and $W_3^{bc} = p(\hat{x})(u(w-L) - x^0)$, $W_3^{TT} = p(\hat{x})u(w-L)$ since it is not worth spending any effort. Hence,

$$\Delta W_3 = W_3^{TT} - W_3^{bc} = p(\hat{x})x^0. \tag{11}$$

Together, TT increases social welfare by the sum of (9), (10) and (11):

$$\Delta W = \Delta W_1 + \Delta W_2 + \Delta W_3 = \hat{x}p(x^0) + x^0 - \int_0^{\hat{x}} p(t)\mathrm{d}t.$$

□

## A.5 Proof of Corollary 3

*Proof.* Due to (4), the technological improvement preserves both the optimal effort $x^0$ and the corresponding loss probability in the benchmark case. But the loss probability becomes lower for every $x$ other than $x^0$. This leads to a larger $A$ and a larger $B_2$ in Figure 1, whereas $B_1 + C$ remains unchanged. Therefore, the overall welfare improvement increases. □

## A.6 Proof of Proposition 4

*Proof.* For the first statement, observe that when $t > \hat{x}$,

$$U^{TT} = w(w-L) < u(w - p(0)L) - 0 < U^0.$$

Electronic copy available at: https://ssrn.com/abstract=3501523

For the second statement, the change of social welfare equals

$$\begin{aligned}\Delta W &= W^{TT} - U^0 \\ &= (1-p(0))u(w) + \int_0^{\hat{x}} [\check{x} - t] f(t)\mathrm{d}t + p(\hat{x})[\check{x} - \hat{x}] \\ &= \check{x} - \int_0^{\hat{x}} p(t)\mathrm{d}t.\end{aligned}$$

The last line follows by applying integration by parts. □

## A.7 Proof of Proposition 3

*Proof.* For the first part of the proposition, observe that once $\tilde{y} = y$ is revealed, the decision problem becomes:

$$\max_x U(x,y) = [1 - p(x,y)]u(w) + p(x,y)u(w-L) - x. \tag{12}$$

The optimal decision $x(y)$ is characterized by the first-order condition:

$$p_1(x(y), y) = -\frac{1}{u(w) - u(w-L)} = -\frac{1}{\hat{x}}.$$

Fully differentiating with respect to $y$ yields

$$\frac{\mathrm{d}x(y)}{\mathrm{d}y} = -\frac{p_{12}(x(y), y)}{p_{11}(x(y), y)},$$

the sign of which is determined by the sign of $\delta'(x)$. For the second part of the proposition, note that $x(y)$ solves (12) for every $y$, so $U(x(y), y) > (=) U(x^0, y)$ whenever $x^0 \neq (=) x(y)$. Under ID and DD, $x(y)$ changes with $y$, so the inequality holds at least for some $y$'s, whereas under CD the equality always holds. This proves the Pareto improvement under ID and DD as well as the absence of welfare effect under CD. To see the change of utilitarian social welfare from $W^0$ to $W^{pTT}$, let $p_u(x) = \int_{\underline{y}}^{\overline{y}} p(x,y)\mathrm{d}G(y)$ be the unbiased prior loss probability and let $x_u^0$ be the optimal benchmark effort given $p_u(x)$. We have:

$$\begin{aligned} W^0 &= [1 - p_u(x^0)]u(w) + p_u(x^0)u(w-L) - x^0 \\ &\leq [1 - p_u(x_u^0)]u(w) + p_u(x_u^0)u(w-L) - x^0 \\ &= \int_{\underline{y}}^{\overline{y}} \left\{ \left[1 - p(x_u^0, y)\right] u(w) + p(x_u^0, y)u(w-L) - x_u^0 \right\} p(x,y)\mathrm{d}G(y) \\ &< \int_{\underline{y}}^{\overline{y}} \left\{ \left[1 - p(x(y), y)\right] u(w) + p(x(y), y)u(w-L) - x(y) \right\} p(x,y)\mathrm{d}G(y) = W^{pTT}. \end{aligned}$$

Under full TT, the posterior distribution of the threshold effort is degenerate. After the improvement of TT, the posterior distribution given $\tilde{y} = y$ is obtained by replacing $p(x)$ by

Electronic copy available at: https://ssrn.com/abstract=3501523

$p(x, y)$ in Proposition **??**. The former is obviously a mean-preserving spread of the latter. According to Gollier (2001), full TT always yields a higher value than an improvement of TT. For (c), consider a simple example where CD is the true case, but the agent misperceives the technology to have ID with respect to $\tilde{y}$. The misperception leads her to reduce (increase) effort when she finds out her risk is higher (lower) than average while she should in fact not change her effort at all. This will make her worse off than if she were correctly informed about CD. Since the improvement of TT does not change welfare under CD, the misperception makes her even worse off than before knowing $y$. For (d), note that if $p(x)$ is biased, that is, $p(x) \neq p_u(x)$, then the second line above is a strict inequality the right hand side of which is the baseline social welfare under the unbiased prior. □

## A.8 Proof of Proposition 5

We first show that any regret-averse decision-maker demands more self-protection than an expected utility maximizer. Inserting $x^0$ to the left hand side of Equation 8 and applying Equation 4 yields:

$$\begin{aligned}
\frac{V'(x^0)}{k} &= f(x^0)g(\hat{x}) - \int_0^{x^0} g'(x^0 - t)f(t)\mathrm{d}t - \int_{x^0}^{\hat{x}} g'(\hat{x} + x^0 - t)f(t)\mathrm{d}t - \left[1 - F(\hat{x})\right] g'(x^0) - F(0)g'(x^0) \\
&= \int_0^{x^0} g'(t)f(x^0)\mathrm{d}t + \int_{x^0}^{\hat{x}} g'(t)f(t)\mathrm{d}t - \int_0^{x^0} g'(t)f(x^0 - t)\mathrm{d}t - \int_{x^0}^{\hat{x}} g'(t)f(\hat{x} + x^0 - t)\mathrm{d}t \\
&\qquad - \left[F(0) + 1 - F(\hat{x})\right] g'(x0)
\end{aligned}$$

The last step follows from integration by subsitution. We may then apply the mean value theorem and obtain the following:

$$\begin{aligned}
\frac{V'(x^0)}{k} &= g'(x_1)\int_0^{x^0} \left[f(x^0) - f(x^0 - t)\right] \mathrm{d}t + g'(x_2)\int_{x^0}^{\hat{x}} \left[f(x^0) - f(\hat{x} + x^0 - t)\right] \mathrm{d}t - \left[F(0) + 1 - F(\hat{x})\right] g'(x^0) \\
&> g'(x_2)\left\{\int_0^{x^0} \left[f(x^0) - f(x^0 - t)\right] \mathrm{d}t + \int_{x^0}^{\hat{x}} \left[f(x^0) - f(\hat{x} + x^0 - t)\right] \mathrm{d}t\right\} \\
&= \left[F(0) + 1 - F(\hat{x})\right] \left[g'(x_2) - g'(x^0)\right] \\
&> 0
\end{aligned}$$

where $0 < x_1 < x^* < x_2 < \hat{x}$. Next, we show that an increase of regret aversion, i.e. an increase of $k$ raises the demand for self-protection. To do this, it suffices to show $V_{xk} > 0$ due to the implicit function theorem, where $V_{xk}$ is the cross derivative of $V$ with respect to $x$ and

Electronic copy available at: https://ssrn.com/abstract=3501523

$k$:

$$\begin{aligned}
V_{xk} &= f(x)g(\hat{x}) - \int_0^x g'(x-t)f(t)\mathrm{d}t - \int_x^{\hat{x}} g'(\hat{x}+x-t)f(t)\mathrm{d}t - [1-F(\hat{x})+F(0)]\,g'(x) \\
&= \frac{V_x + 1 - f(x)\hat{x}}{k} \\
&= \frac{1-f(x)\hat{x}}{k} \\
&> 0.
\end{aligned}$$

The last inequality follows from $f(x) < \frac{1}{\hat{x}}$, which holds because $x > x^0$ and $f(x^0) = 1/\hat{x}$.

Electronic copy available at: https://ssrn.com/abstract=3501523