ISSN 1866-3494

November 30, 2020

ESMT Working Paper 20-02

Human and machine: The impact of machine input on decision-making under cognitive limitations

Tamer Boyaci, ESMT Berlin

Caner Canyakmaz, ESMT Berlin

Francis de Véricourt, ESMT Berlin

Copyright 2020 by ESMT European School of Management and Technology GmbH, Berlin, Germany, https://esmt.berlin/.

All rights reserved. This document may be distributed for free – electronically or in print – in the same formats as it is available on the website of the ESMT (www.esmt.org) for non-commercial purposes. It is not allowed to produce any derivates of it without the written permission of ESMT.

Find more ESMT working papers at ESMT faculty publications, SSRN, RePEc, and EconStor.

Human and Machine: The Impact of Machine Input on Decision-Making Under Cognitive Limitations

Tamer Boyacı, Caner Canyakmaz, Francis de Véricourt
ESMT Berlin, Germany
tamer.boyaci@esmt.org, caner.canyakmaz@esmt.org, francis.devericourt@esmt.org

The rapid adoption of AI technologies by many organizations has recently raised concerns that AI may eventually replace humans in certain tasks. In fact, when used in collaboration, machines can significantly enhance the complementary strengths of humans. Indeed, because of their immense computing power, machines can perform specific tasks with incredible accuracy. In contrast, human decision-makers (DM) are flexible and adaptive but constrained by their limited cognitive capacity. This paper investigates how machine-based predictions may affect the decision process and outcomes of a human DM. We study the impact of these predictions on decision accuracy, the propensity and nature of decision errors as well as the DM's cognitive efforts. To account for both flexibility and limited cognitive capacity, we model the human decision-making process in a rational inattention framework. In this setup, the machine provides the DM with accurate but sometimes incomplete information at no cognitive cost. We fully characterize the impact of machine input on the human decision process in this framework. We show that machine input always improves the overall accuracy of human decisions, but may nonetheless increase the propensity of certain types of errors (such as false positives). The machine can also induce the human to exert more cognitive efforts, even though its input is highly accurate. Interestingly, this happens when the DM is most cognitively constrained, for instance, because of time pressure or multitasking. Synthesizing these results, we pinpoint the decision environments in which human-machine collaboration is likely to be most beneficial.

Key words: machine-learning, rational inattention, human-machine collaboration, cognitive effort

1. Introduction

The increasing adoption of smart machines and data-based technologies have questioned the future role of human-based decisions in organizations (Kleinberg et al. 2017). While new technologies sometimes substitute for labor, a wealth of evidence suggest that they can also complement human skills (see Felten et al. 2019 and references therein). Indeed, the purpose of many real-world applications of supervised machine learning is not to produce a final decision based solely on an algorithm's output, but rather to provide useful information in the form of automated predictions to a human decision-maker (Lipton 2016, Agrawal et al. 2018). Various sectors currently seek to harness such human-machine complementarity, including the defense and health care industries (DARPA 2018), legal and translation services (Katz 2017) or e-sports (OpenAI 2019) among others.

Humans and machines complement each other because machines often substitute for only a subset of the different tasks required to perform an activity (Autor 2015). This is typically the case for

judgment and decision problems. Indeed, human decision-makers rely on their cognitive flexibility to integrate information from vastly diverse sources, including the very context in which these decisions are made (Diamond 2013, Laureiro-Martínez and Brusoni 2018). Machines, by contrast, are much more rigid and can only extract a limited subset of this information (Marcus 2018). Hence, humans may have access to predictive variables that, for example, a machine-learning (ML) algorithm cannot see (Cowgill 2018). However, machine-extracted information can have higher accuracy because of the enormous and reliable quantitative capacity of machines. In contrast, the cognitive capacity of humans is limited, and hence human decision-makers need to constantly balance the quality of their decisions with their cognitive efforts (Payne et al. 1993).

For instance, when deciding on which stocks to invest in, mutual fund managers estimate both idiosyncratic shocks (for stock picking) and aggregate shocks (for market timing) (Kacperczyk et al. 2016). Because of their superior computing capability, ML algorithms identify idiosyncratic shocks with greater success, but fail to detect aggregate ones compared to humans (Fabozzi et al. 2008, Abis 2017). In the medical domain, ML algorithms can easily process large and rich medical histories, but may not obtain valuable information from the personal interaction between physicians and their patients. Similarly, many HR managers base their hiring decisions on information that ML algorithms cannot access (Hoffman et al. 2017).

To the extent that data-based technologies improve the provision of certain information, the co-production of decisions by humans and machines typically boosts the overall quality of these decisions (Mims 2017). For instance, the collaboration between human radiologists and machines improves the overall accuracy of diagnoses for pneumonia over the performances of radiologists alone, or machines alone (Patel et al. 2019). Effective human-machine collaborations such as these[1] are sometimes coined "centaurs" (half-human, half-machine) in the literature and popular press (Case 2018). Yet, the provision of machine-based predictions may not improve all aspects of human decisions. For instance, Stoffel et al. (2018) find that when radiologists take into account the deep-learning analysis of ultrasound images, the diagnoses of breast tumors significantly improve. This is consistent with the claim that human-machine collaboration improves overall performance. However, this improvement mainly stemmed from a radical decrease in false negatives, while the false positive rate did not significantly change.

This impact of machine-based predictions on decision errors, and more generally the time and cognitive efforts that humans put into their decisions, remains largely unknown. As a result, the participation of machines in human decisions may have unintended consequences. Increasing the

[1] The idea of human-machine collaborations –or chess centaurs– were popularized by World Chess Champion Gary Kasparov following his notorious defeat against IBM Deep Blue in 1997. An online chess tournament in 2005 confirmed the superiority of chess centaurs over machines.

number of false positive rates, for instance, may exert undue pressure on a health care delivery system and put healthy patients at risk. And increasing the cognitive load of a decision-maker may slow down the decision process, which may result in delays and congestion.

In this paper, we consider the defining characteristics of human and machine intelligence to address the following fundamental questions: What is the impact of having machine-based predictions on human judgment? In which ways do these predictions influence the decision-making process of humans, the extent of their cognitive efforts, and the nature of their decision errors? In which decision environments are the collaborations between humans and machines more fruitful?

To answer these questions, we consider an elementary decision problem in which an ML algorithm (the machine) assists a human decision-maker (the DM) by assessing part of, but not all, the uncertainty that the DM faces. We model this problem within the theory of rational inattention formalized by Sims (2003, 2006) to capture the most fundamental sources of complementarity between machine and human intelligence. Namely, in our setup, the DM leverages her cognitive flexibility to integrate various sources of information, including her domain knowledge or specific aspects of the context in which the decision is made. Nonetheless, the DM is constrained by her limited cognitive capacity, so that assessing information requires exerting cognitive efforts. The more effort the DM exerts, the more accurate her assessment is. In contrast, the machine does not suffer from this limitation and can provide an accurate assessment of some information at no cost. The machine, however, cannot assess all sources of information, the DM's domain knowledge and the decision context in particular.

The rational inattention framework, within which we develop our model, enables us to represent the DM's cognitive flexibility and limited capacity in a coherent manner. Indeed, this theory assumes that people rationally decide on what piece of information to look for, in what detail, and they do so in an adaptive manner. In particular, the framework endogenously accounts for people's scarce resources, such as time, attention and cognitive capacity as well as the nature of the decision environment. People are free to use any information source, in any order, to generate knowledge at any precision level, but limited cognitive resources lead to information frictions and hence, possible mistaken judgments. In other words, the framework does not impose any a priori restrictions on people's search strategy (cognitive flexibility) other than a limit on the amount of processed information (limited cognitive capacity). More generally, this theory naturally connects the fundamental drivers in human decision-making, such as payoffs, beliefs, and cognitive difficulties in a rational learning setup, and is perceived as a bridging theory between classical and behavioral economics. There is also a growing body of empirical research that finds evidence of decision-making behavior consistent with the theory (Maćkowiak et al. 2018).

In this setup, we analytically compare the DM's choice, error rates, expected payoff, cognitive effort, and overall expected utility when the DM decides alone and when she is assisted by a machine. Our analysis first confirms the aforementioned superiority of the human-machine collaboration. In particular, we show that accuracy and the DM's overall expected utility always (weakly) improve in the presence of a machine. We further find that the machine always reduces false negative errors.

Yet, our results also indicate that machine-based predictions can impair human decisions. Specifically, we find that machine-assisted decisions sometimes *increase* the number of false positives compared to when the DM decides alone. (Incidentally, this finding, along with our result that the machine reduces the false negative rates, offers some theoretical foundation for the empirical results of Stoffel et al. 2018.) In addition, the machine can induce the DM to exert *more* cognitive efforts in expectation, and make her ultimate choice *more uncertain* a priori. In other words, the machine can worsen certain types of decision errors, and increase both the time and variance involved in a decision-making process, which is known to create costly delays and congestion (Alizamir et al. 2013).

We fully characterize the conditions under which these adverse effects occur in our setup. A prominent case is when the DM's prior belief is relatively weak and her cognitive cost of assessing information is relatively high (i.e., her cognitive capacity is reduced due to exogenous time pressure, or consumed by competitive tasks because of multitasking). Yet, those are conditions under which using a machine to offload the DM is most appealing. In other words, improving the efficiency of human decisions by relying on machine-based predictions may in fact backfire precisely when these improvements are most needed. These results hold at least directionally for any decision setting (in terms of payoff and belief structures) and we explain in detail where and why they occur.

The rest of the paper is organized as follows. In §2, we relate our work to the existing literature. In §3, we introduce our basic model of humans and machines and follow in §4 by characterizing the choice behavior and cognitive effort that humans spend, as well as their implied decision errors. In §5, we analyze the impact of machines on these and explain our findings. In §6, we discuss further extensions to the decision and learning environment and investigate their implications for human and machine interaction. Finally, in §7 we present our concluding remarks.

2. Related Literature

Over the past decade, researchers in artificial intelligence have repeatedly demonstrated that algorithmic predictions can match, and at times even outperform, the effectiveness of human decisions in many contexts (see, for instance, Liu et al. 2019 for a recent and systematic review on health care). More recently, however, an emerging literature has focused on improving the collaboration

between machines and humans as opposed to pitching them against each other. For instance, a very recent stream of research in computer science aims at optimizing algorithms by letting them automatically seek human assistance when needed (e.g., Raghu et al. 2019, Wilder et al. 2020, Bansal et al. 2019). More generally, the field aims to improve the interpretability of ML-based predictions so as to facilitate their integration into a human decision-making process (e.g., Doshi-Velez and Kim 2017).

Researchers in management science have also started to study the integration of human judgments into the development of ML algorithms. Ibrahim et al. (2020), for instance, explore how the elicitation process of human forecasts boosts the performance of an algorithm in an experimental setup. Petropoulos et al. (2018) similarly study how human judgment can be used to improve the selection of a forecasting model. Arvan et al. (2019) further provide a literature review on the integration of human judgment into quantitative forecasting methods.

Overall, these different streams of research focus on improving the interaction between humans and machines, either to better train an algorithm or to help humans account for an algorithm's output in their decisions. Not much is known, however, about the impact of machine-based predictions on the human decision-making process.

A few authors have nonetheless analyzed this human-machine interaction in a theoretical decision-making framework. Agrawal et al. (2018) in particular postulate that AI and humans complement one another in that algorithms provide cheap and accurate predictions while humans determine, at a cost, the potential payoffs associated with the decision. Specifically, the authors enrich a standard choice model under uncertainty, in which the DM needs to exert effort to learn her utility function. Our work addresses a different form of complementarity, in which human cognition is flexible but of limited capacity while the machine is rigid but has ample capacity. More recently, Bordt and von Luxburg (2020) propose representing the human-machine joint decision process in a dynamic multi-arm bandit framework. The goal is to study under which conditions humans and machines learn to interact over time and dynamically improve their decisions. In contrast, we study the impact of machine-based predictions on human cognition and decisions. Our setup is therefore static, but it endogenizes the human cognitive efforts.

The rational inattention theory on which our model is based was first introduced by Sims (2003, 2006) and has since been applied in many different contexts, such as discrete choice and pricing (Matějka 2015, Boyacı and Akçay 2018), finance (Kacperczyk et al. 2016) or service systems (Canyakmaz and Boyaci 2020) among many others. Several empirical and experimental studies have further added support to the theory (see, for instance, Bartoš et al. 2016 or Caplin and Dean 2015, and Maćkowiak et al. 2018 for a recent survey). Abis (2017), in particular, proposes an empirical test for a simple model of financial markets made of rationally inattentive humans

and machines with unconstrained capacity. While machines and humans decide independently and may even compete in this setup, our model considers their complementarity.

More generally, our model represents a problem of information acquisition. As such, our work is related to the hypothesis testing Bayesian framework, in which the DM runs a sequence of imperfect tests and dynamically updates her belief accordingly about which decision is best (DeGroot 1970). This approach has been very fruitful for studying a variety of problems, such as the management of research projects or diagnostic services (McCardle et al. 2017, Alizamir et al. 2013, 2019). However, the framework is less suited to represent the cognitive process of a decision-maker. In particular, this Bayesian framework typically assumes that the precision of each test (in the form of false positive and false negative rates) or the order in which they are run are exogenously determined. In contrast, our rational inattention setup does not put any restrictions such as these, and fully endogenizes the level of precision as well as the associated cognitive effort in a tractable way. This enables us to properly account for the flexibility of human cognition (Diamond 2013, Laureiro-Martínez and Brusoni 2018), which is important for our purpose.

Finally, our setup assumes that humans assess information from multiple sources, which jointly designate the true state of the world. This can also be conceptualized as learning states that are characterized by multiple attributes. In this regard, our paper is related to the rich literature on search with multiple attributes (see, for instance, Branco et al. 2012, Olszewski and Wolinsky 2016, Sanjurjo 2017, and references therein). In particular, Huettner et al. (2019) study a multi-attribute discrete choice problem in a rational inattention framework. They retrieve the generalized multinomial logit choice structure in Matějka (2015) for situations where attributes have different information costs. In our model, some attributes are easier to assess when the machine is present, as in Huettner et al. (2019), but we specifically investigate the impact of this on human choice, the extent of decision errors and cognitive efforts.

3. A Model of Human and Machine

In this section, we first present a decision model that captures the flexibility and limited cognitive capacity of the human in a rational inattention framework. We then consider the case where the DM is assisted by a machine.

Consider a human decision-maker (which we will refer to as DM hereon), who needs to correctly assess the true state of the world $\omega \in \{g, b\}$, which can be good ($\omega = g$) or bad ($\omega = b$). We denote by μ the DM's prior belief that the state is good ($\mu = P\{\omega = g\}$). The DM can exert cognitive efforts to evaluate the relevant information and adjust her belief accordingly. The more effort she exerts, the more accurate her evaluation is. When available, a machine-learning algorithm (which we simply refer to as "the machine" in the following) assists the DM by accurately evaluating some

of this information, at no cognitive cost, to account for its immense computing capabilities. Based on her assessment, the DM then announces whether or not the state is good. We denote this choice by $a \in \{y, n\}$ (yes/no), where $a = y$ when the DM chooses the good state and $a = n$ otherwise. The choice is accurate if she chooses $a = y$ and the true state is $\omega = g$, or if $a = n$ and $\omega = b$. The DM enjoys a (normalized) unit of payoff if her decision is accurate, and nothing otherwise. Thus, her expected payoff is the probability that she will make an accurate choice, which we define as the accuracy of her decision. The DM's objective is then to maximize the expected accuracy of her decision,[2] net of any cognitive costs.

3.1. The Human Decision-Maker

The DM is constrained by her limited cognitive capacity, so that assessing available data/information requires exerting cognitive efforts, a process we formalize within the theory of rational inattention. In this framework, the DM is aware of her cognitive limitations and endogenously optimizes how to allocate her effort accordingly. To do this, the DM elicits informative signals about the true state of the world from different sources of information which reduce her prior uncertainty.

Specifically, the DM can elicit any signal $\mathbf{s}$ of any precision level about state $\omega \in \Omega = \{g, b\}$ from any information source. We define an information processing strategy as a joint distribution $f(\mathbf{s}, \omega)$ between signals and states. The DM is free to choose any information processing strategy as long as it is Bayesian consistent with her prior belief (i.e., $\int_{\mathbf{s}} f(\mathbf{s}, g)\, ds = \mu$ must hold). This implies that choosing a strategy $f(\mathbf{s}, \omega)$ is equivalent to determining $f(\omega|\mathbf{s})$, the DM's posterior belief that the true state is w given signal $\mathbf{s}$. In other words, the DM is free to choose the precision of her posterior belief. Thus, the DM may elicit different signals from different information sources in any particular sequence, and make her search for new signals contingent on previous ones to determine the precision of her posterior belief.[3] She may also decide not to process any information at all so that $f(g|\mathbf{s}) = \mu$ or equivalently $f(\mathbf{s}, g) = \mu f(\mathbf{s})$.

Cognitive Effort. The DM's belief about the state of the world specifies the prevalent initial uncertainty. By generating an informative signal $\mathbf{s}$, the DM updates her prior μ to posterior $f(g|\mathbf{s})$. We measure this uncertainty in terms of entropy, which we denote as $H(p)$ for a probability p that the world is in the good state, where $H(p) = -p \log p - (1-p) \log (1-p)$.[4] Entropy is a widely used

[2] In other words, DM's payoffs are the same whether she correctly identifies the good state ($a = y$ when $\omega = g$) or the bad one ($a = n$ when $\omega = b$). This is for the sake of clarity only, though. Our analysis directly extends to a general payoff structure, as we discuss in §6.

[3] Eliciting informative signals can also be imagined as the DM asking a series of yes-or-no questions and observing the outcomes. By choosing an information processing strategy, the DM is effectively choosing what questions to ask and in which sequence.

[4] For a general discrete probability distribution function $P = \{p_\omega; \omega \in \Omega, \sum_{\omega \in \Omega} p_\omega = 1\}$, entropy is defined as $H(P) = \sum_{\omega \in \Omega} p_\omega \log p_\omega$.

measure of uncertainty in the economics literature, because of its different properties and concavity structure in particular, and further corresponds to the expected utility loss from not knowing the state (Frankel and Kamenica 2019).

In our setup, $H(\mu)$ measures the prior level of uncertainty that the DM needs to resolve, and thus fully captures the difficulty level of the decision task. The task presents no difficulty when the DM is fully informed about the state, that is, when $\mu = 1$ or $\mu = 0$ for which $H(\mu)$ is null. The decision task is most difficult when the DM has no prior information about the states, that is, when $\mu = 1/2$ which maximizes $H(\cdot)$. We thus refer to $H(\mu)$ as *the task difficulty* in the following.

Similarly, ex-post entropy $H(f(g|\mathbf{s}))$ measures the level of uncertainty upon eliciting signal $\mathbf{s}$ and thus $\mathbb{E}_{\mathbf{s}}[H(f(g|\mathbf{s}))]$ is the expected level of remaining uncertainty under strategy f, before the DM processes any information. We refer to $\mathbb{E}_{\mathbf{s}}[H(f(g|\mathbf{s}))]$ as *the residual uncertainty* in the following. The expected reduction in uncertainty is then equal to $H(\mu) - \mathbb{E}_{\mathbf{s}}[H(f(g|\mathbf{s}))]$, which corresponds to the mutual information between prior and posterior distributions in information theory and specifies the expected amount of elicited information.[5] This quantity is always positive, that is, information always decreases uncertainty, due to the concavity of entropy $H(\cdot)$.

Reducing uncertainty, however, comes at a cognitive cost. The larger the reduction in uncertainty, the more information is processed and thus the more cognitive effort is required. Following the rational inattention literature, we assume that the DM's cognitive cost is linear in the expected reduction in uncertainty. Formally, the cognitive cost associated with an information processing strategy f is equal to

$$C(f) = \lambda \left(H(\mu) - \mathbb{E}_{\mathbf{s}} \left[H(f(g|\mathbf{s})) \right] \right) \tag{1}$$

where $\lambda > 0$ is the marginal cognitive cost of information which we refer to as *the information cost* in the following.

Overall, information cost λ determines how constrained the DM is in terms of time, attention, and cognitive ability. It may represent the inherent difficulty of assessing a piece of information or the extent to which the DM's cognitive capacity is consumed by competitive tasks, because of time pressure or multitasking. In the latter case, λ is the shadow price of the constraint corresponding to the DM's limited cognitive capacity. Thus, the higher the value of λ, the more effort the DM needs to exert to elicit signals that reduce uncertainty. In the limit where λ is infinite, the DM cannot assess any information and only decides based on her prior belief μ. In contrast, the DM does not have any limit on her capacity when $\lambda = 0$, and can then perfectly assess the true state of the world.

[5] Alternatively, the mutual information can be interpreted as the expected number of questions that the DM needs to ask to implement her information processing strategy according to coding theory (Cover and Thomas 2012). The more questions the DM asks, the more she learns about the true state of the world, which is reflected in the form of a tighter posterior belief.

Decisions and Accuracy. The DM chooses information processing strategy f, at cost $C(f)$, to yield updated belief $f(g|\mathbf{s})$. Given this updated belief, the DM then chooses her action $a \in \{y, n\}$ to maximize accuracy, such that $a = y$ if $f(g|\mathbf{s}) > f(b|\mathbf{s})$ and $a = n$ otherwise (recall that in our setup, the expected payoff is equal to the expected accuracy). Thus, the prior probability that the DM will choose action $a = y$ before she starts assessing any information[6] is equal to

$$p(f) \equiv \int_{\mathbf{s}} \mathbb{I}_{\{f(g|\mathbf{s}) \geq f(b|\mathbf{s})\}} f(\mathbf{s})\, d\mathbf{s},$$

where $\mathbb{I}$ denotes the indicator function, which yields expected accuracy $A(f)$,

$$A(f) \equiv \int_{\mathbf{s}} \max_{a \in \{y,n\}} \{f(g|\mathbf{s})\mathbb{I}_{a=y} + f(b|\mathbf{s})\mathbb{I}_{a=n}\} f(\mathbf{s})\, d\mathbf{s} = \int_{\mathbf{s}} \max\{f(g|\mathbf{s}), f(b|\mathbf{s})\} f(\mathbf{s})\, d\mathbf{s}.$$

The Decision Problem. Anticipating her expected posterior payoff upon receiving signals, the DM first decides on her information acquisition strategy, taking into account the cognitive cost associated with its implementation. The DM then chooses her action. It follows that given her choice of information processing strategy f, the DM enjoys an expected total value of

$$V(f) \equiv A(f) - C(f).$$

She determines her information processing strategy by solving the following optimization problem:

$$\begin{aligned} &\max_{f} V(f) \\ &\text{s.t.} \int_{\mathbf{s}} f(\mathbf{s}, g)\, ds = \mu, \end{aligned} \tag{2}$$

where the constraint guarantees that the DM's information processing strategy is Bayesian consistent with her prior belief.

Given prior μ, we denote by $V^\star(\mu)$, the optimal expected value such that $V^\star(\mu) = V(f^\star)$, where $f^\star$ solves (2). Similarly, we define by $A^\star(\mu)$, $C^\star(\mu)$ and $p^\star(\mu)$ the optimal accuracy, cognitive cost, and choice probability, respectively, given prior μ.

Taken together, our setup captures both the cognitive flexibility and cognitive limitations of humans. In this framework, the DM endogenously decides how to allocate her limited attention and how much effort to put into resolving the prevalent uncertainty. In doing so, the DM chooses how much error she will tolerate and the precision of her decisions. This framework further allows us to account for machine-based predictions in the DM's decision process, as we show next.

[6] Note that the DM commits to a decision with certainty ex-post, i.e., after she assesses the available information. But because the signals she will obtain are unknown before she starts the process, her final decision is random ex-ante.

3.2. Accounting for the Machine

To assess the state of the world, the DM leverages her cognitive flexibility (Diamond 2013, Laureiro-Martínez and Brusoni 2018) to integrate information from diverse sources. The machine, by contrast, only extracts a limited subset of this information (Marcus 2018). Thus, we partition the set of information sources from which signals $\mathbf{s}$ are drawn into two distinct subsets: a first one that both the machine and the DM can evaluate, and a second one which is only available to the DM.

We represent the aggregate information contained in these two subsets as random variables X_1 and X_2, respectively. In particular, r.v. X_2 summarizes the predictive variables that are unobservable to the ML algorithm. These may include information drawn from the DM's domain knowledge or specific aspects of the context in which the decision is made. To put this setup into perspective, consider the medical domain. Random variable X_1 may then represent the statistical summary of all the tangible information that is observable to the algorithm, such as the patient's full medical history. Random variable X_2, on the other hand, may represent the information that the physician obtains through personal interaction with the patient. In contrast to the ML algorithm, the DM can elicit signals from both sources. Recall that we do not impose any restriction on the DM's strategy, particularly the order in which she may assess these sources.

Realization $x_i \in \{-,+\}$ of X_i, $i=1,2$, is such that $x_i = +$ (resp. $x_i = -$) is indicative of a good (resp. bad) state. The true state of the world is good only if all available information is positive,[7] that is, $\omega = g$ if and only if $x_1 = x_2 = +$. We refer to $\pi(x_1,x_2) > 0$ with $(x_1,x_2) \in \{-,+\}^2$ as the DM's prior distribution of (X_1,X_2). Hence, the DM's prior belief that the state is good is equal to $\mu = \pi(+,+)$. In the absence of a machine, the DM needs to allocate her cognitive effort between the assessments of x_1 and x_2.

In contrast to the human, the machine does not suffer from any cognitive limitations due to its virtually unbounded computing capacity. We assume that it can extract the exact value of x_1 at no cognitive cost, so that the DM can dedicate her effort solely to the assessment of x_2. In the presence of the machine, therefore, the DM only assesses x_2 so as to update her new belief, which accounts for the machine's evaluation x_1. Specifically, define μ^x as the DM's new belief that the state is good, given the machine's evaluation $x \in \{-,+\}$. We have, using Bayes' rule with $\mu = \pi(+,+)$,

$$\mu^- = 0 \text{ and } \mu^+ = \frac{\mu}{\mu \; + \; \pi(+,-)} > \mu. \tag{3}$$

That is, a negative evaluation by the machine reveals that the true state is bad, while the DM's belief that the state is good increases with a positive evaluation. It follows from Section 3.1 that when the machine output is x, the optimal expected value, accuracy, cognitive cost, and choice probability, are equal to $V^\star(\mu^x)$, $A^\star(\mu^x)$, $C^\star(\mu^x)$ and $p^\star(\mu^x)$, respectively.

[7] When one positive information suffices to determine the good state, the problem can be made equivalent to the current situation by relabeling the good state and the positive information as the bad and negative ones, respectively.

4. Optimal Decisions, Accuracy and Cognitive Cost

In this section, we characterize optimal choice $p^\star(\cdot)$ as a function of prior belief $\mu \in (0,1)$, from which we deduce the optimal expected value, accuracy, and cognitive cost ($V^\star$, $A^\star$, and $C^\star$, respectively). To that end, we follow Matějka and McKay (2015) who establish that problems of the type (2) where the DM chooses strategy f, are equivalent to problems in which she directly selects the conditional probabilities of choosing action a given state w.[8] The intuition for this equivalence is that a one-to-one correspondence exists between actions a and signals $\mathbf{s}$ in the optimal solution. Indeed, eliciting distinct signals that lead to the same posterior belief (and hence decision) incur additional costs without changing the DM's decision, which is suboptimal. In a discrete choice setting, this yields an optimal solution of GMNL (generalized multinomial logit) form where payoffs include endogenously determined terms. The next Lemma formalizes this result in our setup.

Lemma 1. *Given prior $0 < \mu < 1$, the optimal choice probability $p^*(\mu)$ is the unique solution to the following equations in $p \in [0,1]$,*

$$p = (1-\mu)\, p_b + \mu p_g \tag{4}$$

$$p_g = \frac{p e^{1/\lambda}}{p e^{1/\lambda} + 1 - p} \tag{5}$$

$$p_b = \frac{p}{p + (1-p)\, e^{1/\lambda}}. \tag{6}$$

Further, we have

$$A^\star(\mu) = (1-\mu)\,(1-p_b) + \mu p_g \tag{7}$$

$$C^\star(\mu) = \lambda\left[H(p) - (1-\mu)\, H(p_b) - \mu H(p_g)\right] \tag{8}$$

Probabilities p_g and p_b correspond to the optimal conditional probabilities that the DM chooses y given that the true state is g and b, respectively. Probability p is then the (unconditional) probability of choosing y according to consistency equation (4). Probabilities p_g and p_b also determine the extent of the mistakes the DM tolerates. Specifically, the optimal false positive and false negative rates, which we denote as $\alpha^\star$ and $\beta^\star$, respectively, are equal to

$$\alpha^\star = (1-\mu) p_b \tag{9}$$

$$\beta^\star = \mu(1-p_g) \tag{10}$$

such that $\alpha^\star + \beta^\star = 1 - A^\star$.

[8] Note that this is an "as if" result such that the DM is not actually optimizing over choice probabilities but using an optimal information processing strategy that is behaviorally equivalent to the induced optimal choice probabilities.

4.1. Optimal Decisions

Lemma 1 states that the optimal choice probability $p^\star(\mu)$ corresponding to problem (2) is the solution of a system of equations, which also determines decision accuracy $A^\star(\mu)$, cognitive cost $C^\star(\mu)$, and hence expected value obtained $V^\star(\mu) = A^\star(\mu) - C^\star(\mu)$. The next result provides the explicit solution to these equations.

THEOREM 1. *The optimal choice probability $p^\star(\mu)$ that solves (4), (5) and (6) is equal to*

$$p^\star(\mu) = \begin{cases} 0 & \text{if } \mu \le \underline{\mu} \\ \frac{\mu}{1-e^{-1/\lambda}} - \frac{1-\mu}{e^{1/\lambda}-1} & \text{if } \underline{\mu} < \mu < \overline{\mu} \\ 1 & \text{if } \mu \ge \overline{\mu} \end{cases} \tag{11}$$

where

$$\underline{\mu} = \frac{1}{e^{1/\lambda}+1} < 1/2 < \overline{\mu} = \frac{e^{1/\lambda}}{e^{1/\lambda}+1}.$$

Furthermore, $p^\star(\mu)$ is non-decreasing in μ, $\underline{\mu}$ is increasing in λ and $\bar{\mu}$ is decreasing in λ.

Overall, Theorem 1 characterizes the effect of DM's prior belief μ on her optimal choice probability $p^\star(\mu)$. If the DM's prior belief about the true state of the world is sufficiently strong (i.e., $\mu \ge \overline{\mu}$ or $\mu \le \underline{\mu}$), exerting any effort to learn more about this state is not worth the cognitive cost. The DM then makes an immediate decision without assessing any information, based solely on her prior (i.e., $p^\star(\mu) = 1$ or 0). Otherwise, the DM exerts effort to assess the available information until her belief about the true state of the word is sufficiently strong, at which point she commits to a choice. But, because she does not know what this assessment will reveal a priori, her final decision is uncertain ex-ante (i.e., $0 < p^\star(\mu) < 1$). Furthermore, the stronger the DM believes a priori that the world is in the good state, the more likely she will decide accordingly by choosing $a = y$ (i.e., $p^\star(\mu)$ is non-decreasing in μ).

Theorem 1 also enables characterizing the impact of information cost λ on the optimal choice probability, which we denote by $p^\star(\lambda)$ in the next result with a slight abuse of notation.

COROLLARY 1. *Given prior $0 < \mu < 1$, a positive (possibly infinite) threshold $\bar{\lambda}$ exists such that the optimal choice probability is equal to*

$$p^\star(\lambda) = \begin{cases} \frac{\mu}{1-e^{-1/\lambda}} - \frac{1-\mu}{e^{1/\lambda}-1} & \text{if } \lambda < \bar{\lambda} \\ 0 & \text{if } \lambda \ge \bar{\lambda} \text{ and } \mu < 0.5 \\ 1 & \text{if } \lambda \ge \bar{\lambda} \text{ and } \mu > 0.5, \end{cases} \tag{12}$$

where

$$\bar{\lambda}(\mu) = \left| \log \frac{1-\mu}{\mu} \right|^{-1} \text{ if } \mu \ne 0.5 \text{ and } \bar{\lambda} = +\infty \text{ if } \mu = 0.5.$$

Further, $p^\star(\lambda)$ is decreasing (resp. increasing) in λ, and $\bar{\lambda}$ increasing (resp. decreasing) in μ when $\mu < 0.5$ (resp. $\mu > 0.5$).

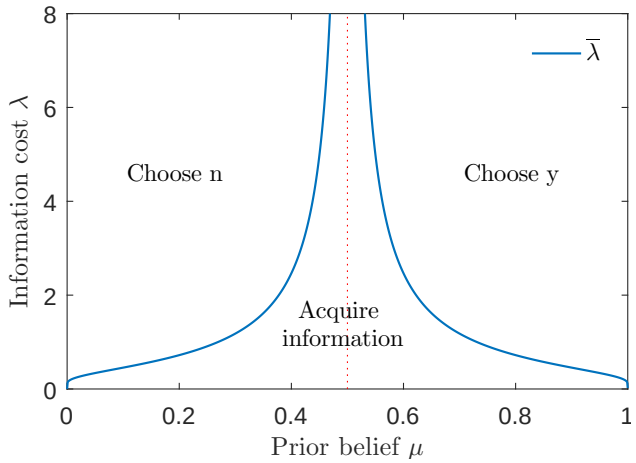


Figure 1 **Effect of prior belief μ on DM's tolerance to information cost $\bar{\lambda}$**

Hence, the DM exerts effort only if the information cost is not too high, that is, less than a threshold. In this case, her probability of choosing the good state increases with the information cost if she favors this state a priori ($\mu > 1/2$), and decreases otherwise. Indeed, the higher the information cost, the less information the DM assesses and thus the less likely her updated belief will significantly change from her prior. Otherwise, she decides a priori that the state is good (resp. bad) if her prior is larger (resp. smaller) than 1/2. In this case, the DM jumps to conclusions as she relies solely on her prior belief without assessing any information. In this sense, threshold $\bar{\lambda}$ determines the DM's tolerance to the information cost. Taken together, Corollary 1 states that the set of prior beliefs for which the DM processes information is an interval centered at 1/2, that shrinks with information cost λ.

Figure 1 depicts the impact of prior μ on threshold $\bar{\lambda}$. When the DM does not have much prior knowledge about the true state of the world (the value of μ is close to 1/2), she is ready to exert a lot of cognitive effort to learn more and hence tolerate high information costs (the value of $\bar{\lambda}$ is high). In particular, the DM always assesses information and exerts effort when the true state is perfectly unknown ($\bar{\lambda} = +\infty$ for $\mu = 1/2$). As the DM is more certain a priori about the true state (μ approaches 0 or 1), she is less willing to exert effort and jumps to conclusions for lower values of information costs ($\bar{\lambda}$ decreases as μ approaches 0 or 1).

4.2. Decision Accuracy and Cognitive Effort

From Lemma 1 and Theorem 1, we obtain $A^\star(\mu)$, $C^\star(\mu)$ and $V^\star(\mu)$ in closed forms, as stated by the following result,

COROLLARY 2. *Given prior μ, $A^\star(\mu)$, $C^\star(\mu)$ and $V^\star(\mu)$ are equal to*

$$A^\star(\mu) = \begin{cases} 1-\mu & \text{if} \quad \mu \leq \underline{\mu} \\ \frac{e^{\frac{1}{\lambda}}}{e^{\frac{1}{\lambda}}+1} & \text{if} \quad \underline{\mu} < \mu < \overline{\mu} \\ \mu & \text{if} \quad \mu \geq \overline{\mu} \end{cases} \tag{13}$$

$$C^{\star}(\mu)=\begin{cases}\lambda\left[H\left(\mu\right)-\varphi\left(\lambda\right)\right] & \text{if} \quad \underline{\mu}<\mu<\overline{\mu}\\ 0 & \text{otherwise}\end{cases} \tag{14}$$

$$V^{\star}\left(\mu\right)=\begin{cases}1-\mu & \text{if} \quad \mu\leq\underline{\mu}\\ \lambda\left[\log\left(e^{\frac{1}{\lambda}}+1\right)-H\left(\mu\right)\right] & \text{if} \quad \underline{\mu}<\mu<\overline{\mu}\\ \mu & \text{if} \quad \mu\geq\overline{\mu}\end{cases} \tag{15}$$

where

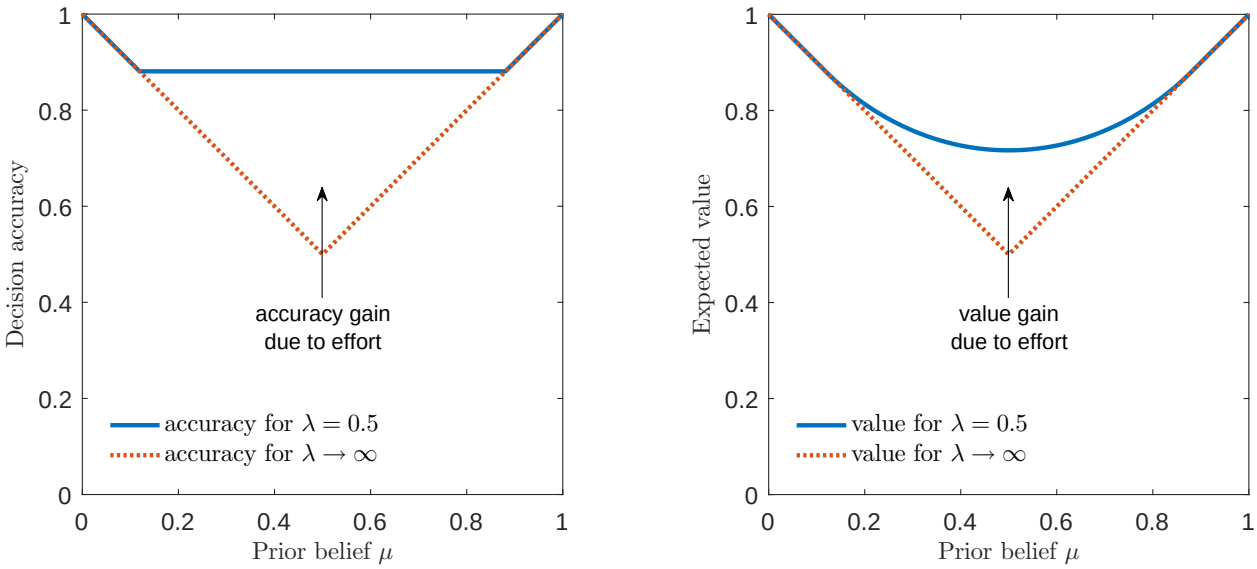
$$\varphi\left(\lambda\right)\equiv\log\left(e^{\frac{1}{\lambda}}+1\right)-\frac{1}{\lambda}\frac{e^{\frac{1}{\lambda}}}{e^{\frac{1}{\lambda}}+1}. \tag{16}$$

Further, $\varphi\left(\lambda\right)$ is increasing in λ, with $\varphi(0)=0$ and $\lim_{\lambda\to\infty}\varphi\left(\lambda\right)=H\left(0.5\right)=\log 2$.

Function $\varphi\left(\lambda\right)$ is the residual uncertainty $\mathbb{E}_{\mathbf{s}}[H(f(g|\mathbf{s}))]$ (see Section 3.1) at optimality. The higher the information cost, the less precise the elicited signals are, and thus the less uncertainty is reduced. Per Corollary 2, residual uncertainty $\varphi\left(\lambda\right)$ is fully determined by the information cost and is independent of the prior. In fact, as long as the DM chooses to process information (i.e., $\underline{\mu}<\mu<\overline{\mu}$), the expected accuracy of her decision depends solely on the information cost and not on her prior belief. Figure 2a illustrates this for a fixed λ. Here, the red dotted curve given by $\max(\mu,1-\mu)$ corresponds to the decision accuracy level the DM obtains when she bases her decision solely on her prior belief (i.e., $\lambda\to\infty$). The solid blue curve is the accuracy function $A(\mu)$ for a finite information cost value, which is constant when the DM chooses to process information. The difference between these two curves precisely corresponds to the gain in accuracy the DM enjoys due to cognitive effort. When the decision task is the most difficult (i.e., when the DM is most uncertain with $\mu=0.5$), the DM obtains the highest accuracy gain, while the magnitude of this gain depends on the information cost.

In contrast, the DM's prior affects expected value $V^{\star}$ through task difficulty $H(\mu)$, if she chooses to exert effort. Specifically, the task difficulty increases the reduction in uncertainty $H(\mu)-\varphi(\lambda)$ that the DM's effort brings about. Thus, Corollary 2 implies that the expected uncertainty reduction and hence the optimal expected cost increase, while the expected value decreases with the task difficulty (i.e., as μ approaches 1/2) which is illustrated in Figure 2b. Similar to Figure 2a, the dotted curve corresponds to the expected value the DM obtains when there is no cognitive effort, in which case it is equal to the expected accuracy. The difference between these two curves gives the value gain the DM enjoys due to her cognitive effort, which is itself the difference between expected value and decision accuracy.

The structure of optimal cost $C^{\star}$ in Corollary 2 sheds further light on thresholds $\underline{\mu}$ and $\bar{\mu}$. Indeed, these thresholds determine when the task difficulty is exactly equal to the optimal reduced uncertainty, that is, $H(\underline{\mu})=H(\bar{\mu})=\varphi(\lambda)$. If $\mu<\underline{\mu}$ or $\mu>\bar{\mu}$, the level of task difficulty is already



(a) Decision accuracy (b) Expected value

Figure 2 DM's accuracy and value functions, and corresponding gains due to cognitive effort

lower than the reduced uncertainty that any cognitive effort would achieve in optimality, that is, $H(\mu) < \varphi(\lambda)$, and the DM prefers to decide a priori, without assessing any information.

That the optimal accuracy is independent of the prior stems from a well-known property of rationally inattentive choice and the fact that the DM maximizes accuracy (net of cognitive costs). Indeed, when some information is processed at optimality, rationally inattentive agents always form the same posterior belief regardless of their prior (see Caplin and Dean 2013). In fact, these optimal posteriors correspond exactly to the belief thresholds that define whether it is economically attractive for the DM to process information ($\underline{\mu}$ and $\overline{\mu}$), which depend only on the payoffs and information cost λ. Intuitively speaking, this means that the DM sharpens her belief by processing costly information, up until the point beyond which it is no longer justified. More specifically, in our context, the DM's optimal posterior belief that the state is good given the aggregate signals that lead to the action $a = y$ (resp. $a = n$) is precisely $\overline{\mu}$ (resp. $\underline{\mu}$) when she processes information. Additionally, since the payoff structure is symmetric in the states, these thresholds (hence, the optimal posteriors) are also symmetric. That is, the DM's posterior belief that the state is good given action $a = y$ (i.e., $\overline{\mu}$) is equal to her posterior belief that state is bad given $a = n$ (i.e., $1 - \underline{\mu}$). In our setup, these are also equal to the accuracy, as it is just the expectation of these over the choice (action) probabilities.

4.3. Decision Errors

Being constrained on cognitive capacity, the decision-maker is bound to make choices based on partial information. Indeed, eliminating all uncertainty is never optimal ($\varphi(\lambda) > 0$ for $\lambda > 0$). This

implies that accuracy is strictly less than one and thus the DM may make false positive and false negative errors, with rates $\alpha^\star$ and $\beta^\star$, respectively. From Theorem 1, we obtain these error rates in closed form in the following corollary.

COROLLARY 3. *Given prior μ, error rates $\alpha^\star(\mu)$ and $\beta^\star(\mu)$ are equal to*

$$\alpha^\star(\mu) = \begin{cases} 0 & \text{if} \quad \mu \le \underline{\mu} \\ 1-\mu & \text{if} \quad \mu \ge \overline{\mu} \\ \frac{\mu\left(e^{1/\lambda}+1\right)-1}{e^{2/\lambda}-1} & \text{otherwise} \end{cases} \tag{17}$$

$$\beta^\star(\mu) = \begin{cases} \mu & \text{if} \quad \mu \le \underline{\mu} \\ 0 & \text{if} \quad \mu \ge \overline{\mu} \\ \frac{e^{1/\lambda}-\mu\left(e^{1/\lambda}+1\right)}{e^{2/\lambda}-1} & \text{otherwise.} \end{cases} \tag{18}$$

If the DM is confident enough that the state is bad ($\mu \le \underline{\mu}$), she chooses $a = n$ without any cognitive effort, preventing her from making a false positive error ($\alpha^\star = 0$) but maximizing her chance of making a false negative one ($\beta^\star = \mu$). The reverse is true ($a = y$, $\alpha^\star = 1-\mu$ and $\beta^\star = 0$) when the DM is sufficiently confident that the state is good ($\mu \ge \bar{\mu}$). Otherwise, DM processes some information and the error rates depend on both the prior and the information cost (with $0 < \alpha^\star < 1-\mu$ and $0 < \beta^\star < \mu$).

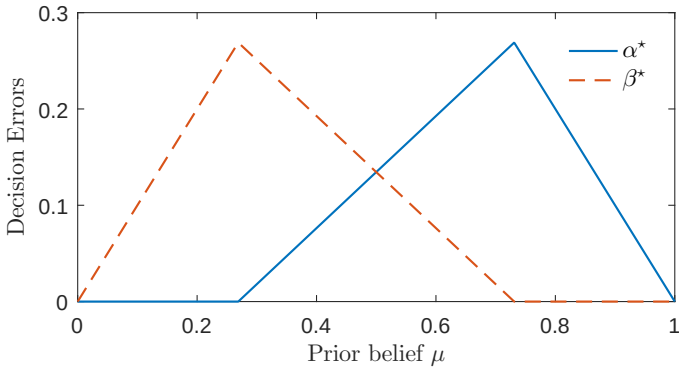


Figure 3 **Error rates $\alpha^\star$ and $\beta^\star$ as a function of prior μ for $\lambda = 1$.**

Figure 3 illustrates the effect of the DM's prior on the error rates. Both $\alpha^\star$ and $\beta^\star$ are piecewise linear and unimodal functions of μ. In particular, when the DM exerts effort ($\underline{\mu} < \mu < \bar{\mu}$), the false positive rate decreases, while the false negative one increases as the prior increases. Note that an increase in prior μ has two conflicting effects on the false positive rate. On one hand, the world is more likely to be in the good state, which decreases the chance of a false positive error. On the other hand, the DM is more likely to choose action $a = y$ for a higher level of μ per Theorem 1, which increases the chance of a false positive error. In essence, Corollary 3 indicates that the second effect always dominates the first one. A similar result holds for the false negative rate.

5. Impact of Machine Input on Human Decisions

Thus far, we have considered a rationally inattentive DM that decides alone. We now investigate how the DM's decision process and its outcomes change when she is assisted by a machine-based assessment. In particular, we compare the DM's decisions, the extent of errors she makes, and the amount of effort she expends with and without the machine.

5.1. Machine-Assisted Decision-Making

In the presence of a machine, the DM first observes the machine's output x_1, which determines her new belief μ^x, $x \in \{+,-\}$, according to (3). The DM then dedicates all her cognitive capacity to evaluating x_2. We denote by $p_m^\star(\mu)$ the resulting ex-ante probability that the DM chooses $a = y$ as a function of her initial prior belief μ. Similarly, $A_m^\star(\mu)$, $C_m^\star(\mu)$, $V_m^\star(\mu)$, $\alpha_m^\star(\mu)$ and $\beta_m^\star(\mu)$ denote decision accuracy, cognitive cost, expected value, and error rates, respectively, that the DM achieves in the presence of the machine. The following (immediate) lemma characterizes these different metrics.

Lemma 2. *Given prior μ, we have*

$$p_m^\star(\mu) = \frac{\mu}{\mu^+} p^\star\left(\mu^+\right), \quad \alpha_m^\star(\mu) = \frac{\mu}{\mu^+}\alpha^\star\left(\mu^+\right) \quad \beta_m^\star(\mu) = \frac{\mu}{\mu^+}\beta^\star\left(\mu^+\right)$$

$$A_m^\star(\mu) = 1 - \frac{\mu}{\mu^+} + \frac{\mu}{\mu^+} A^\star\left(\mu^+\right), \quad C_m^\star(\mu) = \frac{\mu}{\mu^+} C^\star\left(\mu^+\right), \quad V_m^\star = 1 - \frac{\mu}{\mu^+} + \frac{\mu}{\mu^+} V^\star\left(\mu^+\right).$$

Thus, given information cost λ, the decision's outcomes in the presence of the machine can be described with two free parameters $(\mu, \mu^+) \in \mathcal{S} \equiv \{(x,y) \in [0,1]^2, \text{ s.t. } x < y\}$; prior μ, and updated prior μ^+ when the machine gives a positive signal on X_1.

5.2. Impact on Decision Accuracy and Value

Since the machine provides accurate information at no cognitive cost, the machine always improves the expected accuracy and total value of the DM, as stated by the following result.

Proposition 1. *For all $\lambda \geq 0$ and $(\mu, \mu^+) \in \mathcal{S}$, we have $A_m^\star \geq A^\star$ and $V_m^\star \geq V^\star$.*

Figure 2 illustrates Proposition 1. The accuracy levels that can be achieved with a machine for all combinations of $(\mu, \mu^+) \in \mathcal{S}$ correspond to the convex hull of the accuracy curve in Figure 2a (solid blue curve) without the machine. All these points lie above the curve and hence provide greater accuracy. Similarly, the convex hull of the value curve in Figure 2b depicts the set of all possible expected values that the DM can achieve with a machine, showing that it always increases the DM's expected value.

This result provides theoretical support for the growing empirical literature, which shows that human-machine collaborations boost overall accuracy. Interestingly, Proposition 1 is partly driven

by our premise that human cognition is flexible. This feature corresponds in our setup to the unrestricted feasible set of information processing strategies (other than the Bayesian consistency requirement). Indeed, when a priori restrictions are imposed on this feasible set, and hence human cognition is less flexible, accuracy can be shown to sometimes decrease in the presence of the machine.

5.3. Impact on Decisions

The machine improves the expected accuracy and total value of the decision by influencing the DM's choice. The next result determines how the presence of the machine affects this choice as a function of prior μ and posterior belief μ^+.

THEOREM 2. *Given information cost λ, we have*

i) If $\mu^+ \leq \underline{\mu}$, then $p_m^\star = p^\star = 0$.

ii) If $\mu \leq \underline{\mu}$ and $\mu^+ \in \left(\underline{\mu}, \overline{\mu}\right)$, then $p_m^\star > p^\star = 0$.

iii) If $\mu \leq \underline{\mu}$ and $\mu^+ \geq \overline{\mu}$, then $p_m^\star > p^\star = 0$.

iv) If $\underline{\mu} < \mu < \mu^+ < \overline{\mu}$, then $p_m^\star > p^\star$.

v) If $\mu \in \left(\underline{\mu}, \overline{\mu}\right)$ and $\mu^+ \geq \overline{\mu}$, then threshold $\hat{\mu}_c$ exists such that $p_m^\star > p^\star$ if $\mu < \hat{\mu}$ and $p_m^\star \leq p^\star$ otherwise.

vi) If $\mu \geq \overline{\mu}$, then $1 = p^\star > p_m^\star$.

Further, threshold $\hat{\mu}_c$ is equal to

$$\hat{\mu}_c = \left(e^{1/\lambda} + 1 - \frac{e^{1/\lambda} - 1}{\mu^+} \right)^{-1} \geq 1/2 \tag{19}$$

and is decreasing in μ^+.

Overall, Theorem 2 identifies necessary and sufficient conditions under which the presence of the machine *decreases* the DM's probability of choosing $a = y$. This happens when the DM's prior belief is strong enough ($\hat{\mu}_c < \mu$), and a positive assessment by the machine boosts this belief to a sufficiently high level ($\mu^+ \geq \overline{\mu}$). In particular, threshold $\hat{\mu}_c$ is the value of the prior μ, at which the direction of the impact of the machine changes.

Figure 4 illustrates this result in parameter space $\mathcal{S}$, for a given λ. The partition of parameter space $\mathcal{S}$ in six different subsets corresponds to cases *i-vi* in the theorem. Cases i, ii and iii depict situations in which the DM does not exert any effort in the absence of the machine and chooses $a = n$ as a result. This happens when her prior is sufficiently low (i.e., $\mu \leq \underline{\mu}$) per Theorem 1. Similarly, case vi corresponds to situations in which the DM chooses $a = y$ a priori because her prior is sufficiently high (i.e., $\mu \geq \overline{\mu}$). In cases iv and v, however, the DM always exerts effort to assess information in the absence of the machine. The figure demonstrates that threshold $\hat{\mu}_c$ divides

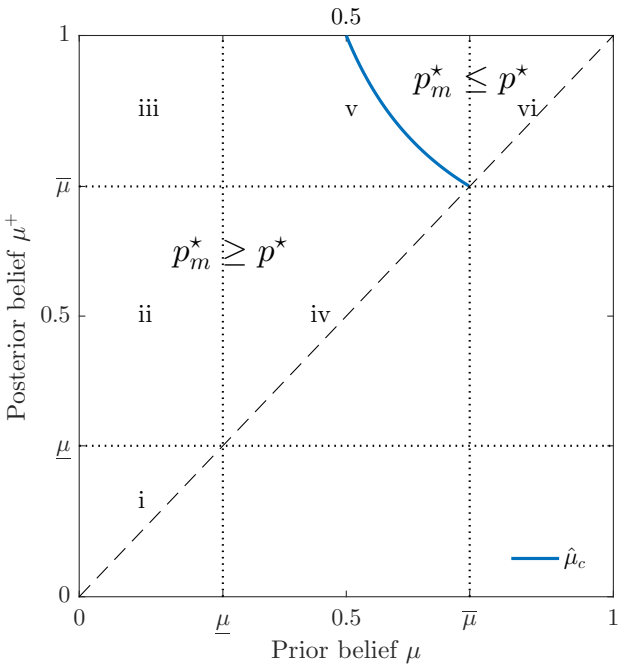


Figure 4 Impact of the machine on DM's decision in parameter space S, for $\lambda = 1$.

space $\mathcal{S}$ into two (top-right and bottom-left) areas, such that the presence of the machine decreases the DM's probability of choosing the good state (i.e., $p^\star_m \leq p^\star$), when (μ, μ^+) lies in the top-right area, and increases the choice probability otherwise.

This result stems from the fact that the machine sometimes dispenses the DM from exerting any effort as well as the impact of the information cost on the DM's choice. To see why, consider the effect of the machine on the DM's choice probability as a function of the information cost, which we characterize next.

Corollary 4. *We have the following:*

- *If $\mu \leq 0.5$, then $p^\star_m \geq p^\star$.*
- *If $\mu > 0.5$, then threshold λ^* exists such that $p^\star_m \geq p^\star$ if $\lambda < \lambda^*$ and $p^\star_m \leq p^\star$ otherwise.*

Further, threshold λ^* is decreasing in prior belief μ with,

$$\lambda^* = \log\left(\frac{\mu^+\mu + \mu - \mu^+}{\mu(1-\mu^+)}\right)^{-1}.$$

In other words, when the DM believes a priori that the good state is more likely ($\mu > 1/2$), the presence of the machine reduces her probability of choosing $a = y$ if the information cost is sufficiently high ($\lambda > \lambda^*$) and increases this probability otherwise. Figure 5 illustrates the result and depicts threshold λ^* as a function of prior μ.

Without the machine, probability $p^\star$ is increasing in the information cost when the DM favors the good state a priori, that is, $\mu > 1/2$ (per Corollary 1). This is because the higher the information

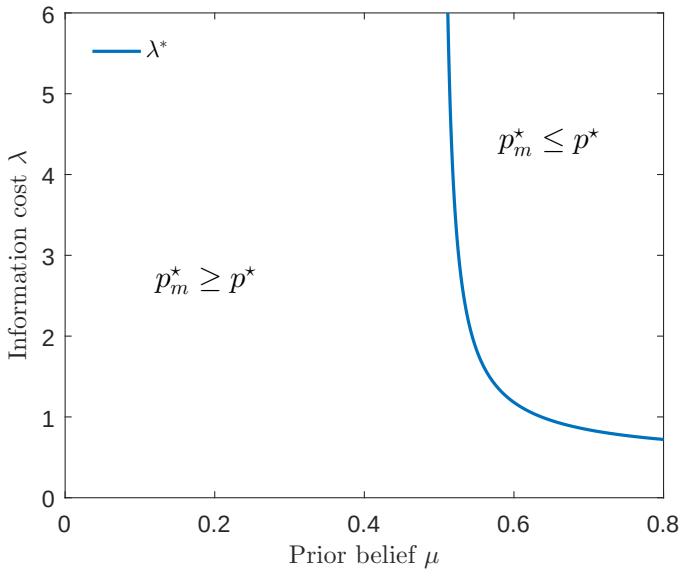


Figure 5 **Impact of the machine on DM's decision as a function of information cost λ and prior μ, for $\mu^+ = 0.8$**

cost, the less information the DM assesses and thus the less likely it is that she will deviate from her prior choice. In the presence of the machine, a positive assessment by the machine boosts the DM's belief, further amplifying this effect. In fact, when information cost λ is greater than threshold $\lambda(\mu^+)$ defined in Corollary 1, a positive assessment by the machine prompts the DM not to exert any additional effort and immediately choose $a = y$ upon receiving the machine's assessment (since $0.5 < \mu < \mu^+$). Thus, probability $p_m^\star$, the ex-ante probability of choosing the good state, corresponds exactly to the chance of a positive result by the machine. And since the machine does not exert any cognitive effort, this probability is independent of the information cost. Hence, probability $p^\star$ increases, while probability $p_m^\star$ remains constant and the former dominates the later when the information cost is sufficiently large.[9]

In other words, a DM without machine sticks to her ex-ante choice with high probability under high information cost. In contrast, a DM assisted by a machine exclusively relies on the machine's result under high information cost. If the machine is not sufficiently likely to confirm the DM's prior, the presence of the machine reduces the DM's chance of choosing the good state. It increases this probability otherwise. In effect, the machine may increases the variability of the DM's decision.

5.4. Impact on Decision Errors

From Proposition 1, we know that the machine always improves accuracy and hence reduces the overall probability of making a mistake. But Theorem 2 indicates that the machine changes the ex-ante probability of choosing an action. This, in turn, should affect the nature of errors that the DM is likely to make. The next result characterizes this effect.

[9] By the same token when $\mu < 1/2$, the choice probability is non-increasing in the information costs which explains why we have $p^\star < p_m^\star$ in this case.

THEOREM 3. *Given information cost λ, $\beta_m^\star \le \beta$ for all $\mu \in [0,1]$. Further, we have*

i) If $\mu^+ \le \underline{\mu}$, then $\alpha_m^\star = \alpha^\star = 0$.

ii) If $\mu \le \underline{\mu}$ and $\mu^+ \in (\underline{\mu}, \overline{\mu})$, then $\alpha_m^\star > \alpha^\star = 0$.

iii) If $\mu \le \underline{\mu}$ and $\mu^+ \ge \overline{\mu}$, then $\alpha_m^\star > \alpha^\star = 0$.

iv) If $\underline{\mu} < \mu < \mu^+ < \overline{\mu}$, and $\mu^+ \in (\underline{\mu}, \overline{\mu})$, then $\alpha_m^\star > \alpha^\star$.

v) If $\mu \in (\underline{\mu}, \overline{\mu})$ and $\mu^+ \ge \overline{\mu}$, then threshold $\hat{\mu}_{fp} < \hat{\mu}_c$ exists such that $\alpha_m^\star > \alpha^\star$ if $\mu < \hat{\mu}_{fp}$, and $\alpha_m^\star \le \alpha^\star$ otherwise.

vi) If $\mu \ge \overline{\mu}$, then $\alpha_m^\star < \alpha^\star$.

Further, threshold $\hat{\mu}_{fp}$ is equal to

$$\hat{\mu}_{fp} = \left(e^{2/\lambda} + e^{1/\lambda} - \frac{e^{2/\lambda} - 1}{\mu^+} \right)^{-1} \tag{20}$$

and is decreasing in μ^+.

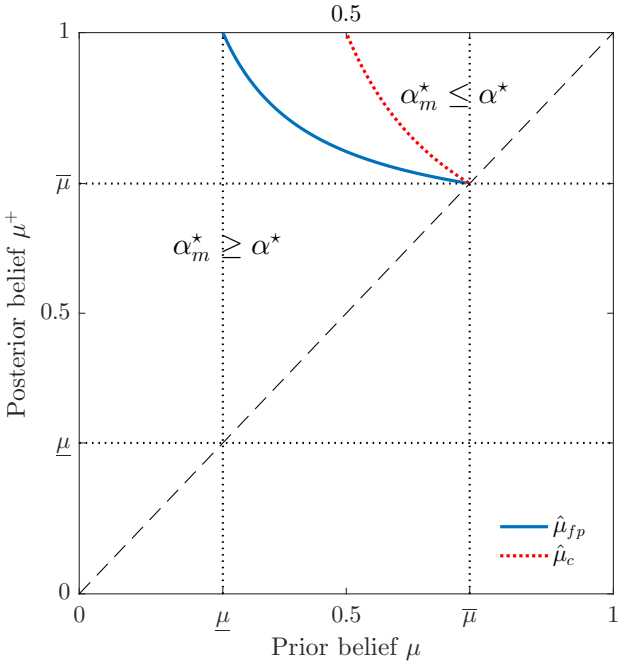


Figure 6 Impact of the machine on DM's false positive error rate in parameter space S, for $\lambda = 1$.

Overall, Theorem 3 states that the machine always improves the false negative rate and thus decreases the DM's propensity of choosing $a = n$ when the state is actually good. This happens even when the machine induces the DM to choose $a = n$ more a priori (i.e., $p_m^\star \le p^\star$ when $\mu \ge \mu_c^+$ per Theorem 2). However, the machine sometimes boosts the false positive rate and thus increases the chance that the DM will choose $a = y$ while the state is actually bad. This happens if the DM's prior belief is not too strong ($\mu < \hat{\mu}_{fp}$). The machine decreases the false positive rate otherwise.

In fact this may happen even when the machine raises the possibility of making this mistake by increasing the overall probability of choosing the good state (i.e., when $\hat{\mu}_{fp} < \mu < \hat{\mu}_c$ per Theorem 2).

Figure 6 illustrates this result in parameter space $\mathcal{S}$, for a given λ. It demonstrates that threshold $\hat{\mu}_{fp}$ divides space $\mathcal{S}$ into two (top-right and bottom-left) areas, such that the presence of the machine decreases the DM's probability of making a false positive type error (i.e., $\alpha_m^\star \leq \alpha^\star$), when (μ, μ^+) lies in the top-right area, and increases otherwise. The effect of information cost λ on DM's error rates, however, is more subtle as the next corollary shows.

COROLLARY 5. *Given prior μ and posterior $\mu^+ > 0.5$, we have*

- *If $\mu \leq \mu^* = 4\mu^+ \frac{1-\mu^+}{\left(2-\mu^+\right)^2}$, then $\alpha_m^\star \geq \alpha^\star$.*
- *If $\mu^* < \mu < 0.5$, there exists two thresholds $\underline{\lambda}_{fp}$ and $\overline{\lambda}_{fp}$ such that $\alpha_m^\star \geq \alpha^\star$ if $\lambda < \underline{\lambda}_{fp}$ and $\lambda > \overline{\lambda}_{fp}$. Otherwise $\alpha_m^\star \leq \alpha^\star$.*
- *If $\mu \geq 0.5$, $\alpha_m^\star \geq \alpha^\star$ if $\lambda < \underline{\lambda}_{fp}$. Otherwise $\alpha_m^\star \leq \alpha^\star$.*

For $\mu^+ \leq 0.5$, we have $\alpha_m^\star \geq \alpha^\star$.

Corollary 5 establishes that regardless of the cost of information, if the DM's prior is sufficiently low, the machine always increases the DM's propensity of making a false positive error which is consistent with Theorem 3. This is because when the DM sufficiently favors the bad state, she chooses $a = n$ more often, which greatly reduces her chance of making a false positive error. In fact, when $\mu < \underline{\mu}$, she never makes a false positive error. On the other hand, a positive machine assessment may render the DM more uncertain (when μ^+ is close to 0.5) or may greatly favor the good state, prompting her to make more false positive errors.

When the DM's prior is too low, the information cost plays a central role in determining the machine's impact on the DM's decision errors. To understand this effect, first consider the case where the DM initially favors the good state (i.e., $\mu > 0.5$). When the information cost is sufficiently low, it is easier for the DM to distinguish the states and less likely that she will make a decision error. However, the machine can increase the DM's chances of making a false positive error by increasing her prior to a sufficiently high level where she chooses $a = y$ directly without acquiring further information. On the other hand, when the information cost is high, the DM without the machine is likely to make a false positive error as she is inclined to choose $a = y$ based on her prior belief (see Corollary 1). The machine, however, can decrease this chance by completely revealing the bad states.

A more subtle effect of the information cost emerges when the DM is sufficiently uncertain, but favors the bad state initially (μ is close but strictly less than 0.5). Again, when the information cost is sufficiently low, she makes fewer false positive errors without the machine as she can still

distinguish the states, and the machine may induce her to choose $a = y$ directly without acquiring further information. However, contrary to the previous case, she also makes fewer false positive errors without the machine when the information cost is sufficiently high, as she is inclined to choose $a = n$ based on her prior belief. Thus, the machine only helps the DM to reduce her false positive errors for moderate information cost levels. Figure 7 illustrates this. The figure plots information cost thresholds $\underline{\lambda}_{fp}$ and $\overline{\lambda}_{fp}$ as functions of prior belief μ for the case where $\mu^+ > 0.5$. The prior belief μ at which the two curves meet precisely corresponds to μ^*. We provide the closed-form characterizations of the two information cost thresholds in the Appendix.

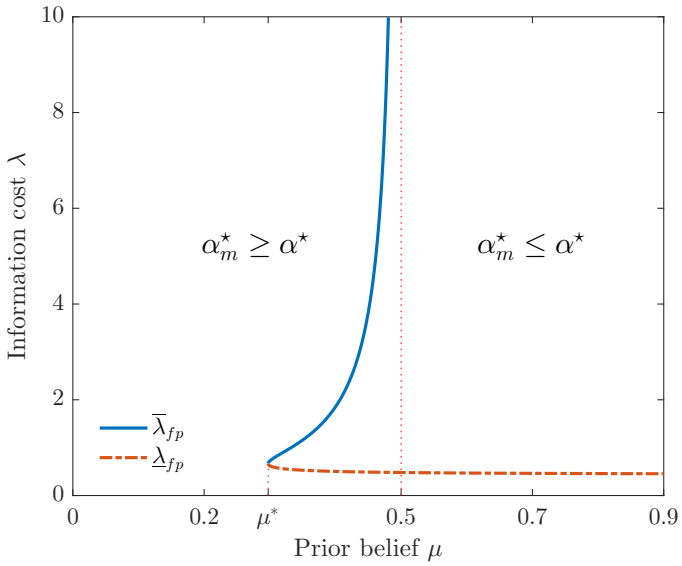


Figure 7 **Impact of the machine on DM's false positive error in information cost λ and prior μ, for $\mu^+ = 0.9$**

5.5. Impact on Cognitive Effort

The machine improves the expected value of human decisions, $V^\star = A^\star - C^\star$, by increasing accuracy $A^\star$ (Proposition 1) due to a decrease in decision errors, but also a change of error types (Theorem 3). An additional and perhaps more intuitive channel by which the machine might improve this expected value is cognitive cost $C^\star$. Indeed, the machine provides information at no cost and may partially relieve the DM of her cognitive effort. This, in turn, should improve the decision's expected value. Yet, the following result, one of our main findings, shows that this is not always the case. In fact, the machine sometimes increases the DM's cognitive cost with $C_m^\star > C^\star$.

THEOREM 4. *Given information cost λ we have,*

i) If $\mu^+ \leq \underline{\mu}$, then $C_m^\star = C^\star = 0$.

ii) If $\mu \leq \underline{\mu}$ and $\mu^+ \in (\underline{\mu}, \overline{\mu})$, then $C_m^\star > C^\star = 0$.

iii) If $\mu \leq \underline{\mu}$ and $\mu^+ \geq \overline{\mu}$, then $C_m^\star = C^\star = 0$.

iv) If $\underline{\mu} < \mu < \mu^+ < \overline{\mu}$*, then threshold* $\hat{\mu}_e \leq 1/2$ *exists such that* $C_m^\star > C^\star$ *if* $\mu < \hat{\mu}_e$ *and* $C_m^\star \leq C^\star$ *otherwise.*

v) If $\mu \in (\underline{\mu}, \overline{\mu})$ *and* $\mu^+ \geq \overline{\mu}$*, then* $0 = C_m^\star < C^\star$.

vi) If $\mu \geq \overline{\mu}$*, then* $C_m^\star = C^\star = 0$.

Furthermore, threshold $\hat{\mu}_e$ *is the unique value of* μ*, for* $\underline{\mu} < \mu < \mu^+ < \overline{\mu}$*, that satisfies*

$$H(\mu) - \frac{\mu}{\mu^+} H(\mu^+) = (1 - \frac{\mu}{\mu^+})\varphi(\lambda) \tag{21}$$

and is decreasing in μ^+.

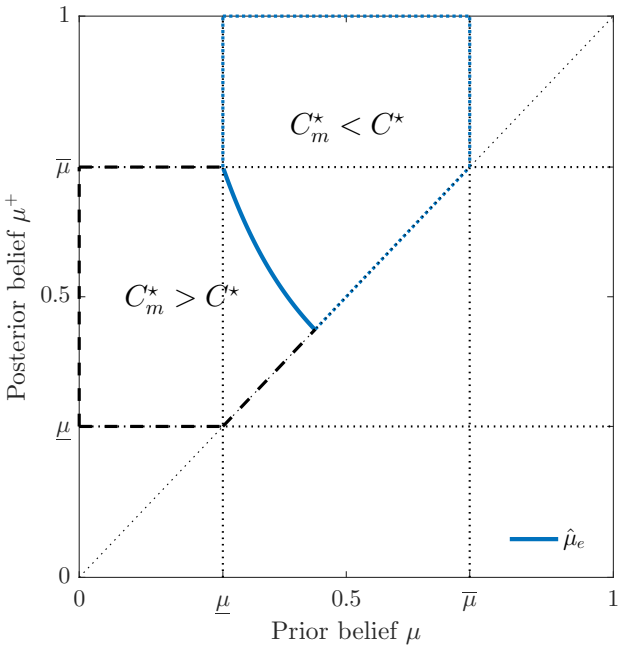


Figure 8 Impact of the machine on DM's cognitive effort in parameter space $\mathcal{S}$, for $\lambda = 1$

Theorem 4 identifies the necessary and sufficient conditions under which the machine induces the DM to exert *more* effort. This happens when the DM sufficiently favors the bad state a priori ($\mu < \hat{\mu}_e \leq 1/2$), which is illustrated in Figure 8. In this case, the task difficulty increases with a positive machine output and the DM needs to exert more effort.

More generally, the machine affects the DM's cognitive cost via the task difficulty and the residual uncertainty ($H(\mu)$ and $\varphi(\lambda)$, respectively, with $C^\star = H(\mu) - \varphi(\lambda)$) but in opposite directions. On one hand, the machine always provides additional information and thus always reduces the task difficulty in expectation ($H(\mu) > \mathbb{E}_{X_1} H(\mu^{X_1})$). This task simplification contributes to reducing the DM's cognitive effort. Note that the effect is ex ante. The DM expects the machine to reduce the difficulty before obtaining the machine assessment. Ex post, a positive result of the machine can increase the task difficulty (i.e., $H(\mu) < H(\mu^+)$). On the other hand, the machine assessment is

precise and hence always decreases the residual uncertainty. In particular, the state is known when the machine's result is negative and, thus, the machine always reduces the residual uncertainty in expectation ($\varphi(\lambda) > P(X_1 = 1)\,\varphi(\lambda)$). This gain in precision contributes to increasing the DM's cognitive effort.

Hence, the machine induces the DM to exert more effort when the precision gain dominates the task simplification that the machine brings about. This happens when the prior is sufficiently small and the information cost is large enough, as stated by the following corollary.

COROLLARY 6. *If $\mu^+ \geq 0.5$ and $\mu > 1-\mu^+$, then $C_m^\star \leq C^\star$. Otherwise, a unique threshold λ_e^* exists such that $C_m^\star > C^\star$ if $\lambda > \lambda_e^*$ and $C_m^\star \leq C^\star$ otherwise. Furthermore, threshold λ_e^* satisfies*

$$\frac{H(\mu) - \frac{\mu}{\mu^+} H(\mu^+)}{1 - \frac{\mu}{\mu^+}} = \varphi(\lambda_e^*) \tag{22}$$

and is increasing in prior belief μ.

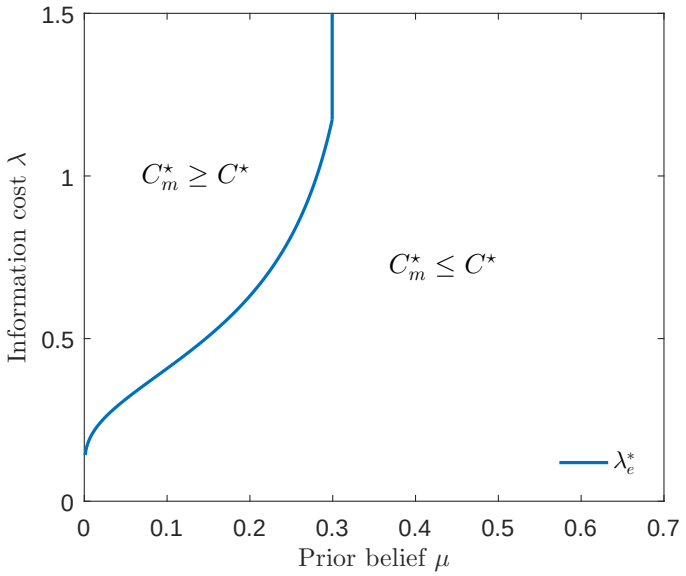


Figure 9 **Impact of the machine on DM's cognitive effort in information cost λ and prior μ, for $\mu^+ = 0.7$**

In other words, if the DM sufficiently believes that the state is good ($\mu > 1-\mu^+$), the machine always decreases her cognitive costs in expectation. Otherwise, the machine increases the cognitive cost when the information cost is sufficiently large ($\lambda > \lambda_e^*$). This means, perhaps surprisingly, that a machine induces more cognitive efforts when the DM is not sure about the good state and is already experiencing a high level of cognitive load (i.e., for a high λ), but reduces these efforts when she is relatively sure about the good state *or* has already ample cognitive capacity (i.e., for a low λ). Figure 9 illustrates this. The figure depicts λ_e^* as a function of prior belief μ for the case where $\mu^+ = 0.7$. Note that λ_e^* is defined only for belief values that are less than $1-\mu^+ = 0.3$ and determines whether the machine increases the DM's cognitive effort or not.

6. Generalized Payoffs

Our base model assumes that the DM's payoff corresponds to the overall accuracy of her decisions. Accuracy is indeed the main performance metric of interest in the empirical literature on machine-assisted decisions. However, our framework can also account for a general payoff structure of the form $u(a,\omega)$, for $(a,\omega) \in \{y,n\} \times \{g,b\}$. An analysis similar to our basic case can show that our results and insights continue to hold in this more general setup. Nonetheless, this general payoff structure creates an asymmetry in the DM's incentives that our previous analysis does not capture. Specifically, a DM who cares only about accuracy does not prefer one state over the other. By contrast, an asymmetric payoff structure may induce the DM to allocate more effort toward a specific state at the expense of the other. This has implications for her choices and decision errors. For instance, if identifying the bad state is more important (as is perhaps the case in a medical setting where it corresponds to a sick patient), the DM may tolerate false negatives more and choose $a = n$ more often.

More specifically, we can normalize any payoff structure $u(a,\omega)$ such that $u(y,g) = 1$ and $u(n,g) = 0$ without loss of generality (see Appendix). To avoid any trivial solution, we assume that $u(n,b) > u(y,b)$ (otherwise, the payoff of $a = y$ dominates the payoff of $a = n$ in all states of the world and the DM directly chooses the former without processing information). In this setup, difference $\delta = u(n,b) - u(y,b)$ denotes the net value of correctly identifying the bad state. (The net value of correctly identifying the good state is always equal to one.) Thus, the DM prefers to correctly identify the bad state over the good state if and only if $\delta > 1$. In our base model, $\delta = 1$ with $u(n,b) = 1$ and $u(y,b) = 0$, so that the DM is indifferent between identifying the good and the bad states.

Figure 10 depicts the impact of the machine on the DM's decision (analogous to Figure 4) for $\delta < 1$ and $\delta > 1$. The figures demonstrate that the structure of our result continues to hold for more general payoffs. In addition, the figure reveals that the set of values of beliefs μ and μ^+ for which $p_m^\star \geq p^\star$ widens as δ increases. Indeed, increasing δ decreases the likelihood that the DM will choose $a = y$ as this option becomes a less attractive alternative. Accordingly, the threshold level $\overline{\mu}$ on the DM's prior belief that warrants immediate ex-ante $a = y$ decision increases. That is, the DM needs to be more confident about the good state to choose $a = y$ without the need to spend further cognitive effort. According to Theorem 2, we already know that the machine induces the DM to choose $a = y$ when her posterior is less than $\overline{\mu}$.

Similar results hold for the false positive error rate and expected cognitive effort that the DM exerts. The set of prior values for which the machine induces fewer false positives ($\alpha_m^\star \leq \alpha^\star$) and reduces cognitive effort ($C_m^\star \leq C^\star$) shrinks as δ increases (see Appendix). And as in our base case, the machine consistently reduces the false negative rate regardless of the incentive structure across states.

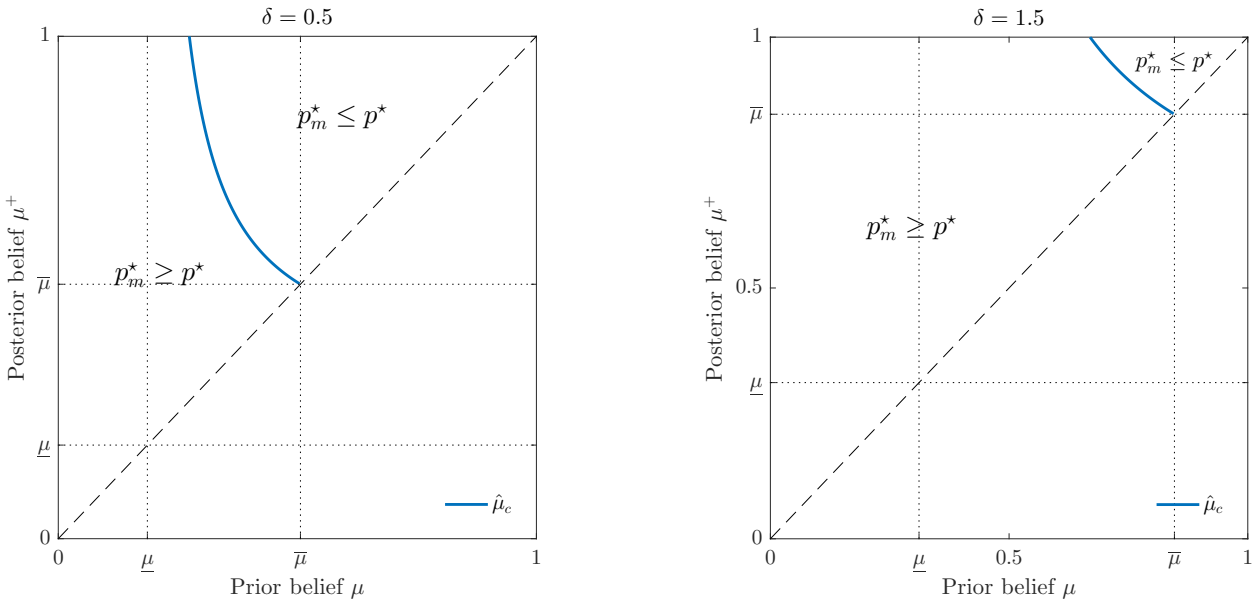


Figure 10 **Impact of incentive structures on DM's decision ($\lambda = 1$)**

7. Concluding Remarks

Humans have always been interested in harnessing technology and machine capabilities for competitive advantage. With the advent of data-based technologies and AI, the collaboration between humans and machine has moved even more to the forefront. This stems from the increasing recognition that human and machines can complement each other in performing tasks and making decisions. In this paper, we develop an analytical model to study the impact of such collaborations on human judgment and decision-making. Our model incorporates the quintessential distinguishing features of human and machine intelligence in a primary decision-making setting under uncertainty: the flexibility of humans to attend to information from diverse sources (and, in particular, the human domain knowledge and the decision context), but under limited cognitive capacity, and in contrast, the rigidity of machines that only process a limited subset of this information, but with great efficiency and accuracy.

We integrate these features endogenously utilizing the rational inattention framework, and analytically characterize the decisions as well as the cognitive effort spent. Comparing the case when the human decides alone to the case with machine input, we are able to discern the impact of machine-based predictions on decisions and expected payoff, accuracy, error rates, and cognitive effort. To put these results in perspective, consider a generic medical assessment setup, in which machine-based predictions (e.g., ML algorithm processing digital images) provide diagnostic input to the physician. The physician can conduct more assessments and tests with the patient. When *both* assessments are positive, then the patient is "sick." The prior reflects the true nature of the *disease's incidence* within the patient population (probability of patient being sick).

Our findings suggest that the machine improves overall diagnostic accuracy (Proposition 1) by decreasing the number of misdiagnosed sick patients (Theorem 3). The machine further boosts the physician's propensity to diagnose patients as healthy when the disease's incidence is high (Theorem 1), and to misdiagnose healthy patients more often when the incidence is low. The physician also exerts less cognitive efforts with the machine, when the disease's incidence is high (Theorem 4). In contrast, the machine induces the physician to exert more cognitive effort when the disease's incidence is low and the physician is under significant time pressure (Corollary 6).

In this example, the patient is sick when both assessments are positive, which corresponds to our basic setup. Other information structures, however, are possible. For instance, consider a generic judicial ruling task, in which machine-based predictions (e.g., ML algorithm checking evidence authenticity, or lie-detection test) provide evidence to the judge. The judge can analyze additional data relevant to the case. When *any* assessment is positive, then the suspect is "guilty." The prior reflects the true nature of the *crime level* within the suspect population (probability of suspect being guilty). As we briefly mention in Section 3.2, our basic setup can account for this situation by relabeling the good state and the positive information in our model as the bad and negative ones, respectively. This also reverses the effect in our results, as Table 1 depicts. This table provides a flavor of the different implications that could arise from our findings in two hypothetical settings fitting to our context.

Medical assessment & diagnostic accuracy	Judicial ruling & conviction accuracy
• Overall diagnostic accuracy is improved	• Overall conviction accuracy is improved
• Fewer misdiagnosed sick patients	• Fewer acquitted guilty suspects
• More patients declared healthy when the disease incidence is high	• More suspects declared guilty when crime level is low
• More misdiagnosed healthy patients when the disease incidence is low	• More convicted non-guilty suspects when crime level is high
• Physician spends less cognitive effort to diagnose when the incidence is high	• Judge spends less cognitive effort to assess evidence when crime level is low
• Physician spends more cognitive effort to diagnose when the incidence is low and time is constrained	• Judge spends more cognitive effort to assess evidence when the crime level is high and time is constrained

Table 1: **Impact of the machine on human decisions for two generic settings**

As the above examples highlight, the incorporation of machine-based predictions on human decisions is not always beneficial, neither in terms of the reduction of errors nor the amount of cognitive effort. The theoretical results we present underscore the critical impact machine-based predictions have on human judgment and decisions. Our analysis also provides prescriptive guidance on

when and how machine input should be considered, and hence on the design of human-machine collaboration. We offer both hope and caution.

On the positive side, we establish that, on average, accuracy improves due to this collaboration. However, this comes at the cost of making certain decision errors more and increased cognitive effort, in particular when the prior belief (on the "good" state) is relatively weak. Consequently, applications of machine-assisted decision-making is certainly beneficial when there is a priori sufficient confidence in the good state to be identified. In this case, the machine input has a tendency toward "confirming the rather expected," and this provably decreases all error rates and improves the "efficiency" of the human by reducing cognitive effort. In sharp contrast, caution is advised for applications that involve searching and identifying a somewhat unlikely good state, especially when the human is significantly constrained in cognitive capacity due to limited time or multitasking. In this case, a positive indication by the machine has a strong effect of "falsifying the expected." The resulting increase in task difficulty not only deteriorates the efficiency of the human by inducing more cognitive effort, but also increases her propensity to incorrectly conclude that the state is good. Hence, human-machine collaboration may fail to provide the expected efficiency gain (and to some extent accuracy) precisely when they are arguably most desirable.

As a final remark, this paper focuses on tasks for which human cognitive flexibility complements machine accuracy. In particular, the machine only processes a subset of the relevant information. If the situation is reversed and the DM's domain knowledge, for instance, is a subset of the information that the machine has access to, this complementarity is mute. This corresponds to tasks for which machine predictions can substitute for and even outperform human decision-making.

References

Abis S (2017) Man vs. machine: Quantitative and discretionary equity management, working paper, Columbia University.

Agrawal AK, Gans JS, Goldfarb A (2018) Prediction, judgment and complexity: A theory of decision making and artificial intelligence. Technical report, National Bureau of Economic Research.

Alizamir S, de Véricourt F, Sun P (2013) Diagnostic accuracy under congestion. *Management Science* 59(1):157–171.

Alizamir S, de Véricourt F, Sun P (2019) Search under accumulated pressure. *Operations Research* .

Arvan M, Fahimnia B, Reisi M, Siemsen E (2019) Integrating human judgement into quantitative forecasting methods: A review. *Omega* 86:237–252.

Autor DH (2015) Why are there still so many jobs? the history and future of workplace automation. *The Journal of Economic Perspectives* 29(3):3–30.

Bansal G, Nushi B, Kamar E, Weld DS, Lasecki WS, Horvitz E (2019) Updates in human-ai teams: Understanding and addressing the performance/compatibility tradeoff. *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 2429–2437.

Bartoš V, Bauer M, Chytilová J, Matějka F (2016) Attention discrimination: Theory and field experiments with monitoring information acquisition. *American Economic Review* 106(6):1437–75.

Bordt S, von Luxburg U (2020) When humans and machines make joint decisions: A non-symmetric bandit model. *arXiv preprint arXiv:2007.04800* .

Boyacı T, Akçay Y (2018) Pricing when customers have limited attention. *Management Science* 64(7):2995–3014.

Branco F, Sun M, Villas-Boas JM (2012) Optimal search for product information. *Management Science* 58(11):2037–2056.

Canyakmaz C, Boyaci T (2020) Queueing systems with rationally inattentive customers. *ESMT Berlin Working Paper* .

Caplin A, Dean M (2013) Behavioral implications of rational inattention with shannon entropy. Technical report, National Bureau of Economic Research.

Caplin A, Dean M (2015) Revealed preference, rational inattention, and costly information acquisition. *American Economic Review* 105(7):2183–2203.

Case N (2018) How to become a centaur. *Journal of Design and Science* .

Cover TM, Thomas JA (2012) *Elements of information theory* (John Wiley & Sons, Hoboken, NJ).

Cowgill B (2018) The impact of algorithms on judicial discretion: Evidence from regression discontinuities. Technical report, Technical Report. Working paper.

DARPA M (2018) Darpa announces $2 billion campaign to develop next wave of ai technologies. URL `www.darpa.mil/news-events/2018-09-07`, accessed: 2019-09-20.

DeGroot M (1970) *Optimal Statistical Decisions* (McGraw-Hill, New York).

Diamond A (2013) Executive functions. *Annual review of psychology* 64:135–168.

Doshi-Velez F, Kim B (2017) Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608* .

Fabozzi FJ, Focardi SM, Jonas CL (2008) On the challenges in quantitative equity management. *Quantitative Finance* 8(7):649–665.

Felten EW, Raj M, Seamans R (2019) The variable impact of artificial intelligence on labor: The role of complementary skills and technologies. *Available at SSRN 3368605* .

Frankel A, Kamenica E (2019) Quantifying information and uncertainty. *American Economic Review* 109(10):3650–80.

Hoffman M, Kahn LB, Li D (2017) Discretion in hiring. *The Quarterly Journal of Economics* 133(2):765–800.

Huettner F, Boyacı T, Akçay Y (2019) Consumer choice under limited attention when alternatives have different information costs. *Operations Research* 67(3):671–699.

Ibrahim R, Kim SH, Tong J (2020) Eliciting human judgment for prediction algorithms. *Available at SSRN* .

Kacperczyk M, Van Nieuwerburgh S, Veldkamp L (2016) A rational theory of mutual funds' attention allocation. *Econometrica* 84(2):571–626.

Katz M (2017) Welcome to the era of the ai coworkera. URL `www.wired.com/story/welcome-to-the-era-of-the-ai-coworker/`, accessed: 2019-09-20.

Kleinberg J, Lakkaraju H, Leskovec J, Ludwig J, Mullainathan S (2017) Human decisions and machine predictions. *The quarterly journal of economics* 133(1):237–293.

Laureiro-Martínez D, Brusoni S (2018) Cognitive flexibility and adaptive decision-making: Evidence from a laboratory study of expert decision makers. *Strategic Management Journal* 39(4):1031–1058.

Lipton ZC (2016) The mythos of model interpretability. *arXiv preprint arXiv:1606.03490* .

Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, Mahendiran T, Moraes G, Shamdas M, Kern C, et al. (2019) A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The lancet digital health* 1(6):e271–e297.

Maćkowiak B, Matejka F, Wiederholt M (2018) Rational inattention: A disciplined behavioral model. *Goethe University Frankfurt mimeo* .

Marcus G (2018) Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631* .

Matějka F (2015) Rigid pricing and rationally inattentive consumer. *Journal of Economic Theory* 158:656–678.

Matějka F, McKay A (2015) Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review* 105(1):272–98.

McCardle KF, Tsetlin I, Winkler RL (2017) When to abandon a research project and search for a new one. *Operations Research* .

Mims C (2017) Without Humans, Artificial Intelligence Is Still Pretty Stupid. *Wall Street Journal* ISSN 0099-9660.

Olszewski W, Wolinsky A (2016) Search for an object with two attributes. *Journal of Economic Theory* 161:145–160.

OpenAI (2019) Openai five defeats dota 2 world champions. URL `openai.com/blog/openai-five-defeats-dota-2-world-champions/`, accessed: 2020-08-20.

Patel BN, Rosenberg L, Willcox G, Baltaxe D, Lyons M, Irvin J, Rajpurkar P, Amrhein T, Gupta R, Halabi S, et al. (2019) Human–machine partnership with artificial intelligence for chest radiograph diagnosis. *NPJ digital medicine* 2(1):1–10.

Payne JW, Payne JW, Bettman JR, Johnson EJ (1993) *The adaptive decision maker* (Cambridge university press).

Petropoulos F, Kourentzes N, Nikolopoulos K, Siemsen E (2018) Judgmental selection of forecasting models. *Journal of Operations Management* 60:34–46.

Raghu M, Blumer K, Corrado G, Kleinberg J, Obermeyer Z, Mullainathan S (2019) The algorithmic automation problem: Prediction, triage, and human effort. *arXiv preprint arXiv:1903.12220* .

Sanjurjo A (2017) Search with multiple attributes: Theory and empirics. *Games and Economic Behavior* 104:535–562.

Sims CA (2003) Implications of rational inattention. *Journal of monetary Economics* 50(3):665–690.

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

Stoffel E, Becker AS, Wurnig MC, Marcon M, Ghafoor S, Berger N, Boss A (2018) Distinction between phyllodes tumor and fibroadenoma in breast ultrasound using deep learning image analysis. *European Journal of Radiology Open* 5:165–170, URL `http://dx.doi.org/10.1016/j.ejro.2018.09.002`.

Wilder B, Horvitz E, Kamar E (2020) Learning to complement humans. *arXiv preprint arXiv:2005.00582* .

Appendix A: Proofs of Results

Proof of Lemma 1 (5) and (6) follow from Theorem 1 in Matějka and McKay (2015) which are obtained for action $a=y$. $A^\star$ and $C^\star$ are by definition.

Proof of Theorem 1 By (4), (5) and (6), we have $p=(1-\mu)\frac{p}{p+(1-p)e^{1/\lambda}}+\mu\frac{pe^{1/\lambda}}{pe^{1/\lambda}+1-p}$ which gives

$$\overline{p}=\frac{\mu}{1-e^{-1/\lambda}}-\frac{1-\mu}{e^{1/\lambda}-1}.$$

Choice probability is 1 when $\overline{p}\geq 1$, or equivalently, $\mu\geq\overline{\mu}=\frac{e^{1/\lambda}}{e^{1/\lambda}+1}$. Similarly, choice probability is 0 when $\overline{p}\leq 0$ or equivalently $\mu\leq\underline{\mu}=\frac{1}{e^{1/\lambda}+1}$. As $\overline{p}$ is linearly increasing in μ with first order derivative $\frac{1}{1-e^{-1/\lambda}}+\frac{1}{e^{1/\lambda}-1}>0$, $p^\star$ is increasing in μ. Finally, $\overline{\mu}$ is decreasing in λ as $\frac{d\overline{\mu}}{d\lambda}=-\frac{1}{\lambda^2}e\frac{\frac{1}{\lambda}}{\left(e^{\frac{1}{\lambda}}+1\right)^2}<0$ and $\underline{\mu}$ is increasing in λ as $\frac{d\underline{\mu}}{d\lambda}=\frac{1}{\lambda^2}\frac{e^{\frac{1}{\lambda}}}{\left(e^{\frac{1}{\lambda}}+1\right)^2}>0$.

Proof of Corollary 1 $\mu\leq\underline{\mu}=\frac{1}{e^{1/\lambda}+1}\Leftrightarrow\left(\log\frac{1-\mu}{\mu}\right)^{-1}\leq\lambda$ which yields an $a=n$ decision by Theorem 1. Note that $\log\frac{1-\mu}{\mu}$ is positive when $\mu<0.5$. Similarly, $\mu\geq\overline{\mu}=\frac{e^{1/\lambda}}{e^{1/\lambda}+1}\Leftrightarrow\lambda\geq\left(\log\frac{\mu}{1-\mu}\right)^{-1}$ which leads to $a=y$ decision and the log term is positive when $\mu>0.5$. Therefore, for any $\mu\neq 0.5$, $\overline{\lambda}$ can be written in absolute terms, that is, $\overline{\lambda}=\left|\log\frac{1-\mu}{\mu}\right|^{-1}$. Then (12) follows. Furthermore, $\frac{d}{d\lambda}p^\star(\lambda)=\frac{1}{\lambda^2}\frac{e^{-\frac{1}{\lambda}}}{\left(e^{-\frac{1}{\lambda}}-1\right)^2}(2\mu-1)$ which is positive when $\mu>0.5$ and negative when $\mu<0.5$. Hence the monotonicity result follows.

Proof of Corollary 2 We can write the DM's accuracy in (7) in terms of optimal posterior beliefs that she constructs as

$$A^\star = (1-\mu)(1-p_b) + \mu p_g = \gamma(b|n)(1-p^\star) + \gamma(g|y)p^\star$$

where $\gamma(\omega|a)$ denotes the optimal posterior that the state is ω given action a. When $\mu < \underline{\mu}$, $A^\star = 1-\mu$ as $p^\star = 0$ and $\gamma(b|n) = 1-\mu$. Similarly, when $\mu > \overline{\mu}$, $A^\star = \mu$ as $p^\star = 1$ and $\gamma(g|y) = \mu$. For the case where $\mu \in \left[\underline{\mu}, \overline{\mu}\right]$, we use the optimal posterior characterizations that are given in Lemma 3 (in Appendix C) for $\delta = 1$, which yields $\gamma(g|y) = \overline{\mu} = \frac{e^{1/\lambda}}{e^{1/\lambda}+1}$ and $\gamma(g|n) = \underline{\mu} = \frac{1}{e^{1/\lambda}+1}$. Note that $\gamma(g|y) = \gamma(b|n)$ and we have $A^\star = \overline{\mu} = \frac{e^{1/\lambda}}{e^{1/\lambda}+1}$.

Using the symmetry of mutual information (see Cover and Thomas 2012), we can write (8) as

$$C^\star = \lambda\left[H(\mu) - pH(\gamma(g|y)) - (1-p)H(\gamma(g|n))\right].$$

Assume that $\mu \in \left[\underline{\mu}, \overline{\mu}\right]$. Then, as $\gamma(g|y) = 1-\gamma(g|n)$ we have $H(\gamma(g|y)) = H(\gamma(g|n))$ from the symmetry of the entropy function H in $[0,1]$. Then, $C^\star$ becomes

$$\begin{aligned} C^\star &= \lambda\left[H(\mu) - H\left(\frac{e^{1/\lambda}}{e^{1/\lambda}+1}\right)\right] = \lambda\left[H(\mu) + \frac{e^{1/\lambda}}{e^{1/\lambda}+1}\log\frac{e^{1/\lambda}}{e^{1/\lambda}+1} + \frac{1}{e^{1/\lambda}+1}\log\frac{1}{e^{1/\lambda}+1}\right] \\ &= \lambda\left[H(\mu) + \frac{1}{\lambda}\frac{e^{1/\lambda}}{e^{1/\lambda}+1} - \log\left(e^{1/\lambda}+1\right)\right]. \end{aligned}$$

When, $\mu \notin \left[\underline{\mu}, \overline{\mu}\right]$, $C^\star = 0$, as $\gamma(g|y) = \gamma(g|n) = \mu$. Finally, $V^\star$ is found by taking the difference $A^\star - C^\star$.

Proof of Corollary 3 When $\mu \le \underline{\mu}$, $p_b = p_g = 0$, hence $\alpha^\star = 0$ and $\beta^\star = \mu$ by (9) and (10). Similarly, when $\mu > \overline{\mu}$, $p_b = p_g = 1$, hence $\alpha^\star = 1-\mu$ and $\beta^\star = 0$. Now assume $\mu \in \left[\underline{\mu}, \overline{\mu}\right]$. Writing (9) and (10) in terms of optimal posteriors in Lemma 3 for $\delta = 1$ and plugging in optimal choice in (11), we obtain

$$\begin{aligned} \alpha^\star &= (1-\mu)p_b = (1-\gamma(g|y))p = \frac{1}{e^{1/\lambda}+1}\left(\frac{\mu}{1-e^{-1/\lambda}} - \frac{1-\mu}{e^{1/\lambda}-1}\right) = \frac{\mu\left(e^{1/\lambda}+1\right)-1}{e^{2/\lambda}-1} \\ \beta^\star &= \mu(1-p_g) = \gamma(g|n)(1-p) = \frac{1}{e^{1/\lambda}+1}\left(1 - \frac{\mu}{1-e^{-1/\lambda}} + \frac{1-\mu}{e^{1/\lambda}-1}\right) = \frac{e^{1/\lambda} - \mu\left(e^{1/\lambda}+1\right)}{e^{2/\lambda}-1}. \end{aligned}$$

Proof of Lemma 2 Using $\mu^- = 0$, the result follows by (11), (17), (18), (13), (14) and (15) for $p^\star$, $\alpha^\star$, $\beta^\star$, $A^\star$, $C^\star$, $V^\star$, respectively.

Proof of Proposition 1 Accuracy $A^\star$ in (13) is convex in μ for $[0,1]$. Then, by Jensen's inequality,

$$A^\star(\mu) = A^\star\left(\left(1-\frac{\mu}{\mu}\right)\mu^- + \frac{\mu}{\mu^+}\mu^+\right) \le \left(1-\frac{\mu}{\mu^+}\right)A^\star(\mu^-) + \frac{\mu}{\mu^+}A^\star\left(\mu^+\right) = 1 - \frac{\mu}{\mu^+} + \frac{\mu}{\mu^+}A^\star\left(\mu^+\right) = A_m^\star(\mu).$$

Similarly, the value function (15) is also convex in μ. To see this, note first that $V(\mu)$ is linearly decreasing in $\left[0, \underline{\mu}\right]$, convex in $\left[\underline{\mu}, \overline{\mu}\right]$ (since the entropy function H is concave) and linearly increasing in $[\overline{\mu}, 1]$. Furthermore, the slope of $\lambda\left[\log\left(e^{\frac{1}{\lambda}}+1\right) - H(\mu)\right]$ is the same at both of these cutoff points. More specifically,

$$\frac{d}{d\mu}_{|\mu=\underline{\mu}} \lambda\left[\log\left(e^{\frac{1}{\lambda}}+1\right) - H(\mu)\right] = \lambda\log\frac{\mu}{1-\mu}_{|\mu=\underline{\mu}} = \lambda\log\frac{\frac{1}{e^{1/\lambda}+1}}{1-\frac{1}{e^{1/\lambda}+1}} = -1.$$

Similarly, $\lambda\log\frac{\overline{\mu}}{1-\overline{\mu}} = 1$. Since the slope is increasing in μ, $V(\mu)$ is convex in μ for $\mu \in [0,1]$. By Jensen's inequality $V^\star(\mu) \le V_m^\star(\mu)$.

Proof of Theorem 2 Note that *i*), *ii*), *iii*) and *vi*) follow by the optimal choice probability in (11) and the fact that $p_m^\star = \mu/\mu^+ p^\star(\mu^+)$.

iv) Using (11) and $\frac{\mu}{\mu^+} < 1$, we have

$$p_m^\star = \frac{\mu}{\mu^+}\left(\frac{\mu^+}{1-e^{-1/\lambda}} - \frac{1-\mu^+}{e^{1/\lambda}-1}\right) = \frac{\mu}{1-e^{-1/\lambda}} - \frac{\frac{\mu}{\mu^+}-\mu}{e^{1/\lambda}-1} > \frac{\mu}{1-e^{-1/\lambda}} - \frac{1-\mu}{e^{1/\lambda}-1} = p^\star$$

v) $p_m^\star > p^\star \ \frac{\mu}{\mu^+} > \frac{\mu}{1-e^{-1/\lambda}} - \frac{1-\mu}{e^{1/\lambda}-1}$ which can equivalently be written as

$$\frac{1}{e^{1/\lambda}-1} > \mu^+\left(\frac{1}{1-e^{-1/\lambda}} + \frac{1}{e^{1/\lambda}-1} - \frac{1}{\mu^+}\right). \tag{23}$$

The right hand side is always positive since $\mu^+ > \frac{e^{1/\lambda}}{e^{1/\lambda}+1}$. That is,

$$\frac{1}{1-e^{-1/\lambda}} + \frac{1}{e^{1/\lambda}-1} - \frac{1}{\mu^+} < 0 \Leftrightarrow \mu^+ > \frac{e^{1/\lambda}-1}{e^{1/\lambda}+1}$$

which is always true since $\mu^+ > \frac{e^{1/\lambda}}{e^{1/\lambda}+1}$. Then, (23) can be written as

$$\mu^+ < \frac{\frac{1}{e^{1/\lambda}-1}}{\frac{1}{1-e^{-1/\lambda}} + \frac{1}{e^{1/\lambda}-1} - \frac{1}{x}} = \left(e^{1/\lambda} + 1 - \frac{e^{1/\lambda}-1}{\mu^+}\right)^{-1} = \widehat{\mu}_c.$$

Note that $\widehat{\mu}_c$ is decreasing in μ^+ and for $\mu^+ = 1$, $\widehat{\mu}_c = 0.5$. Then $\widehat{\mu}_c \geq 0.5$.

Proof of Corollary 4 Assume $\mu \in \left[\underline{\mu}, \overline{\mu}\right]$ for a fixed λ. By Theorem 2, since $\widehat{\mu}_c \geq 0.5$, $p_m^\star \geq p^\star$ for $\mu \leq 0.5$. For $\mu < \underline{\mu}$, $p_m^\star \geq p^\star$ by *i*, *ii* and *iii*. This proves the first part. When $\mu > 0.5$, $p_m^\star \leq p^\star$ if $\mu \geq \widehat{\mu}_c$ for a fixed λ. Using (19), we have

$$\mu \geq \widehat{\mu}_c \Leftrightarrow \frac{1}{\mu} \leq e^{1/\lambda} + 1 - \frac{e^{1/\lambda}-1}{\mu^+} \Leftrightarrow \frac{\mu^+ + 1 - \frac{\mu^+}{\mu}}{1-\mu^+} \geq e^{1/\lambda} \Leftrightarrow \lambda \geq \left(\log \frac{\mu^+ + 1 - \frac{\mu^+}{\mu}}{1-\mu^+}\right)^{-1} = \lambda^*.$$

Proof of Theorem 3 *i*), *ii*), *iii*) and *vi*) follow directly by the optimal error probability functions $\alpha^\star(\mu)$ and $\beta^\star(\mu)$ in (17) and (18), respectively and the fact that $\alpha_m^\star = \mu/\mu^+ \alpha^\star(\mu^+)$ and $\beta_m^\star = \mu/\mu^{+\beta^\star}(\mu^+)$.

iv) Using (17), (18) and $\frac{\mu}{\mu^+} < 1$ we have

$$\alpha_m^\star = \frac{\mu}{\mu^+}\alpha^\star\left(\mu^+\right) = \frac{\mu\left(e^{1/\lambda}+1\right) - \frac{\mu}{\mu^+}}{e^{2/\lambda}-1} > \frac{\mu\left(e^{1/\lambda}+1\right)-1}{e^{2/\lambda}-1} = \alpha^\star$$

and

$$\beta_m^\star = \frac{\mu}{\mu^+}\beta^\star\left(\mu^+\right) = \frac{\frac{\mu}{\mu^+}e^{1/\lambda} - \mu\left(e^{1/\lambda}+1\right)}{e^{2/\lambda}-1} < \frac{e^{1/\lambda} - \mu\left(e^{1/\lambda}+1\right)}{e^{2/\lambda}-1} = \beta^\star.$$

v) $\alpha_m^\star > \alpha^\star$ when

$$\begin{aligned}
\frac{\mu}{\mu^+}\left(1-\mu^+\right) > \frac{\mu\left(e^{1/\lambda}+1\right)-1}{e^{2/\lambda}-1} &\Leftrightarrow \mu\left(\frac{1}{\mu^+}-1\right) > \mu\frac{1}{e^{1/\lambda}-1} - \frac{1}{e^{2/\lambda}-1} \\
&\Leftrightarrow \mu\left(\frac{1}{e^{1/\lambda}-1} - \frac{1}{\mu^+} + 1\right) < \frac{1}{e^{2/\lambda}-1} \\
&\Leftrightarrow \mu < \frac{\frac{1}{e^{2/\lambda}-1}}{\frac{1}{e^{1/\lambda}-1} - \frac{1}{\mu^+} + 1} = \left(e^{2/\lambda} + e^{1/\lambda} - \frac{e^{2/\lambda}-1}{\mu^+}\right)^{-1} = \widehat{\mu}_{fp}.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\widehat{\mu}_{fp} = \left(e^{2/\lambda} + e^{1/\lambda} - \frac{e^{2/\lambda}-1}{\mu^+}\right)^{-1} &< \left(e^{1/\lambda} + 1 - \frac{e^{1/\lambda}-1}{\mu^+}\right)^{-1} = \widehat{\mu}_c \\
\Leftrightarrow e^{2/\lambda} - 1 > \frac{e^{2/\lambda}-1}{\mu^+} - \frac{e^{1/\lambda}-1}{\mu^+} &\Leftrightarrow e^{1/\lambda} + 1 > \frac{e^{1/\lambda}}{\mu^+} \Leftrightarrow \mu^+ > \frac{e^{1/\lambda}}{e^{1/\lambda}+1} = \overline{\mu}
\end{aligned}$$

which is always true by assumption. For the false negative, since $\beta^\star(\mu^+) = 0$, we have $\beta_m^\star = 0 < \beta^\star$.

Proof of Corollary 5 Solving the belief threshold $\hat{\mu}_{fp}$ in (20) for λ, we obtain the following two roots;

$$\underline{\lambda}_1 = \frac{1}{\log\left[\frac{1}{2}\left(\frac{1}{1-\mu^+} - 1 + \sqrt{\frac{\mu(2-\mu^+)^2 - 4(1-\mu^+)\mu^+}{\mu(1-\mu^+)^2}}\right)\right]}$$
$$\overline{\lambda}_1 = \frac{1}{\log\left[\frac{1}{2}\left(\frac{1}{1-\mu^+} - 1 - \sqrt{\frac{\mu(2-\mu^+)^2 - 4(1-\mu^+)\mu^+}{\mu(1-\mu^+)^2}}\right)\right]}$$

Note that these roots are real valued when expression inside the square root is positive, that is, when $\mu > 4\mu^+\frac{1-\mu^+}{(2-\mu^+)^2}$. Otherwise there are no real roots and $\alpha_m^\star \geq \alpha^\star$. Assume that this condition holds and consider $\overline{\lambda}_1$. It is positive when

$$\frac{1}{2}\left(\frac{1}{1-\mu^+} - 1 - \sqrt{\frac{\mu(2-\mu^+)^2 - 4(1-\mu^+)\mu^+}{\mu(1-\mu^+)^2}}\right) > 1 \Leftrightarrow \frac{1}{1-\mu^+} - 3 > \sqrt{\frac{\mu(2-\mu^+)^2 - 4(1-\mu^+)\mu^+}{\mu(1-\mu^+)^2}}.$$

Note that first $\mu^+ > 2/3$ should hold so that the left hand side is positive. Then, it can be shown that (after some elementary mathematical operations) $\mu < 1/2$ should hold as well. Similarly, it can be shown that for $\underline{\lambda}_1$ is positive, either when $\mu^+ > 2/3$ or when both $\mu^+ < 2/3$ and $\mu > 1/2$ are satisfied. This means that since $\mu < \mu^+$ by default, when $\mu^+ < 1/2$, there are no positive real-valued roots, and hence $\alpha_m^\star \geq \alpha^\star$. Let us define $\overline{\lambda}_{fp} = \overline{\lambda}_1^+$ and $\underline{\lambda}_{fp} = \underline{\lambda}_1^+$ where $x^+ = \max\{0, x\}$. Assume $\mu^+ > 2/3$. Then when $\mu < 1/2$, $\overline{\lambda}_{fp} = \overline{\lambda}_1$ and $\underline{\lambda}_{fp} = \underline{\lambda}_1$. Taking the first order derivative of the belief threshold $\hat{\mu}_{fp}$ in (20) with respect to λ, we see that it is positive when $\mu^+ - 2e^{1/x}(1-\mu^+) > 0$, that is, when the two roots $\overline{\lambda}_1$ and $\underline{\lambda}_1$ exist, $\hat{\mu}_{fp}$ is first decreasing than increasing. Then, since $\alpha_m^\star \geq \alpha^\star$ when $\mu \leq \hat{\mu}_{fp}$, it is true also when $\lambda \leq \underline{\lambda}_{fp}$ or $\lambda \geq \overline{\lambda}_{fp}$. Assume now that $1/2 < \mu^+ < 2/3$. Then, when $\mu < 1/2$, $\overline{\lambda}_{fp} = \underline{\lambda}_{fp} = 0$, that is, $\alpha_m^\star \geq \alpha^\star$ for $\lambda > \overline{\lambda}_{fp}$. When $\mu > 1/2$, $\overline{\lambda}_{fp} = 0$ and $\underline{\lambda}_{fp} = \underline{\lambda}_1$, and $\alpha_m^\star \geq \alpha^\star$ for $\lambda < \underline{\lambda}_{fp}$.

Proof of Theorem 4 *i*, *ii*, *iii*, *v* and *vi* correspond to cases where either DM's prior belief μ or posterior belief μ^+ induces DM to spend no cognitive effort. In this case, total cognitive cost is zero in at last one of the cases and the results follow. For case *iv* where the DM processes information in both of these cases, $C_m^\star > C^\star$ when $\frac{\mu}{\mu^+}\left(H\left(\mu^+\right) - \varphi\left(\lambda\right)\right) > H\left(\mu\right) - \varphi\left(\lambda\right)$. Note that the left hand side is a positive increasing function of μ while the right hand side is a concave function that takes its maximum at $\mu = 0.5$. At $\mu = \underline{\mu}$, right hand side is zero and left hand side is positive. At the other extreme when $\mu = \mu^+$, both sides are equal. This means that for $\mu^+ \geq 0.5$, the two functions cross at a single point between $\left(\underline{\mu}, \mu^+\right)$. For $\mu^+ < 0.5$, both functions are increasing. Hence, they cross only if slope of the right hand side function at μ^+ is less than the slope of the left hand side function. The slopes are equal when

$$\frac{H\left(\mu^+\right) - \varphi\left(\lambda\right)}{\mu^+} = \log\frac{1-\mu^+}{\mu^+}\mu^+ \Leftrightarrow \mu^+ = 1 - e^{-\varphi(\lambda)} = \widehat{\mu}_e^+.$$

Hence, for $\mu^+ \leq \widehat{\mu}_e^+ < 0.5$, we have $C_m^\star \geq C^\star$ for all $\mu < \mu^+$. When $\mu^+ \in (\widehat{\mu}_e^+, \overline{\mu})$, the unique threshold $\widehat{\mu}_e$ satisfies

$$\frac{\widehat{\mu}_e}{\mu^+}\left(H\left(\mu^+\right) - \varphi\left(\lambda\right)\right) = H\left(\widehat{\mu}_e\right) - \varphi\left(\lambda\right). \tag{24}$$

Furthermore, left hand side of (24) is decreasing in μ^+ since

$$\frac{H\left(\mu^+\right)}{\mu^+} = \frac{\mu^+\log\frac{1-\mu^+}{\mu^+} - H\left(\mu^+\right)}{\left(\mu^+\right)^2} = \frac{\mu^+\log\frac{1-\mu^+}{\mu^+} + \mu^+\log\mu^+ + (1-\mu^+)\log\left(1-\mu^+\right)}{\left(\mu^+\right)^2} = \frac{\log\left(1-\mu^+\right)}{\left(\mu^+\right)^2} < 0.$$

Therefore the crossing point that satisfies (24) and hence $\widehat{\mu}_e$ is decreasing in μ^+. Lastly, as the concave right hand side function in (24) takes its maximum at 0.5, the crossing point is less than that point, i.e., $\widehat{\mu}_e \leq 0.5$.

Proof of Corollary 6 Assume $\mu^+ \geq 0.5$ and $\mu > 1-\mu^+$. Then $H(\mu) > H(\mu^+)$ since H is symmetric around $\mu = 0.5$. Then

$$C_m^\star = \frac{\mu}{\mu^+}\left(H\left(\mu^+\right) - \varphi(\lambda)\right) < \frac{\mu}{\mu^+}\left(H(\mu) - \varphi(\lambda)\right) < H(\mu) - \varphi(\lambda) = C^\star.$$

Assume otherwise. Then $C_m^\star > C^\star$ if $\frac{H(\mu) - \frac{\mu}{\mu^+}H\left(\mu^+\right)}{1-\frac{\mu}{\mu^+}} < \varphi(\lambda)$. We show that the left hand side is increasing in μ. To see this take the first order derivative;

$$\frac{d}{d\mu}\frac{H(\mu) - \frac{\mu}{\mu^+}H(\mu^+)}{\left(1-\frac{\mu}{\mu^+}\right)} = \frac{\left(\log\frac{1-\mu}{\mu} - \frac{H\left(\mu^+\right)}{\mu^+}\right)\left(1-\frac{\mu}{\mu^+}\right) + \frac{1}{\mu^+}\left(H(\mu) - \frac{\mu}{\mu^+}H(\mu^+)\right)}{\left(1-\frac{\mu}{\mu^+}\right)^2}.$$

Simplifying the numerator, we have

$$\begin{aligned}
&\left(1-\frac{\mu}{\mu^+}\right)\log\frac{1-\mu}{\mu} - \frac{H(\mu^+)}{\mu^+} + \frac{H(\mu)}{\mu^+} = \left(\mu^+ - \mu\right)\log\frac{1-\mu}{\mu} + H(\mu) - H\left(\mu^+\right)\\
&= \left(\mu^+ - \mu\right)\log(1-\mu) - \left(\mu^+ - \mu\right)\log\mu - \mu\log\mu - (1-\mu)\log(1-\mu) - H\left(\mu^+\right)\\
&= -\left(1-\mu^+\right)\log(1-\mu) - \mu^+\log\mu - H\left(\mu^+\right).
\end{aligned}$$

This is decreasing in μ as the first order derivative is $\frac{\mu-\mu^+}{1-\mu} < 0$. Evaluating at $\mu = \mu^+$ (which is the largest possible μ) we obtain zero, that is, $-(1-\mu^+)\log(1-\mu^+) - \mu^+\log\mu^+ - H(\mu^+) = 0$. This means the first order derivative of the left hand side is positive. Note also that the right hand side is increasing in λ with $\lim_{\lambda\to\infty}\varphi(\lambda) = \log 2$ while the left hand side is constant. To see this, take the first order derivative $\varphi'(\lambda) = \frac{1}{\lambda^3}\frac{e^{\frac{1}{\lambda}}}{\left(e^{\frac{1}{\lambda}}+1\right)^2} > 0$. Now, when μ is at its maximum, $\mu = \mu^+$, we have

$$\frac{H(1-\mu^+) - \frac{\mu}{\mu^+}H(\mu^+)}{1-\frac{\mu}{\mu^+}} = \frac{\left(1-\frac{\mu}{\mu^+}\right)H(\mu^+)}{1-\frac{\mu}{\mu^+}} = H\left(\mu^+\right) < \varphi(\lambda).$$

The maximum value entropy function H can take is $\log 2$ and for $\mu < \mu^+$, the maximum value that the left hand side can get is less than $\log 2$. Then this means there exists a unique λ that satisfies (22).

Appendix B: General Payoff Structure

B.1. Normalizing the Payoffs

One can transform any general payoff matrix $\widehat{u}(a,\omega)$ with $a \in \{y,n\}$ and $\omega \in \{g,b\}$ by first subtracting $\widehat{u}(n,g)$ from each payoff, and then scaling each by $1/\left(\widehat{u}(y,g) - \widehat{u}(n,g)\right)$. Then, the new payoff structure becomes

$$\begin{aligned}
u(n,g) &= \widehat{u}(n,g) - \widehat{u}(n,g) = 0 \qquad & u(y,g) &= \frac{\widehat{u}(y,g) - \widehat{u}(n,g)}{\widehat{u}(y,g) - \widehat{u}(n,g)} = 1\\
u(y,b) &= \frac{\widehat{u}(y,b) - \widehat{u}(n,g)}{\widehat{u}(y,g) - \widehat{u}(n,g)} = a \qquad & u(n,b) &= \frac{\widehat{u}(n,b) - \widehat{u}(n,g)}{\widehat{u}(y,g) - \widehat{u}(n,g)} = c.
\end{aligned}$$

Information cost parameter λ, should then be scaled by $1/\left(u(y,g) - u(n,g)\right)$, to arrive at an identical behavioral structure. That is, the new information cost should be $\lambda' = \frac{\lambda}{u(y,g)-u(n,g)}$. The reason is that subtracting $\widehat{u}(n,g)$ from each payoff does not change the DM's problem since the payoff differences (i.e., incentives) stay the same. Therefore, there is no need to change λ. However scaling each payoff by a constant also scales the differences between them which creates a different incentive structure. To avoid this, one needs to scale the information cost also by the same constant.

B.2. Impact of Machine on DM's Decision Errors and Cognitive Costs for General Payoffs

When DM's incentives change, the machine's impact on the extent of errors that the DM makes does not structurally change. In particular, as in our baseline model, when the machine assists the DM with some accurate information, the DM's false negative error always decreases as it completely eliminates the possibility of bad state in some cases. Similarly, the machine can increase the DM's propensity to make false positive errors in some cases. In particular, there still exists a unique threshold $\hat{\mu}_{fp}$ on the DM's prior belief that determines whether the DM makes more or fewer false positive errors with the machine. Furthermore, the larger the net value of correctly identifying bad state δ, the larger the parameter space where the DM makes more false positive errors with the machine. This is because the region where the DM is inclined to choose $a = y$ more with the machine is larger (see Figure 10). The effect of δ on DM's propensity to make false a positive error is illustrated in Figure 11.

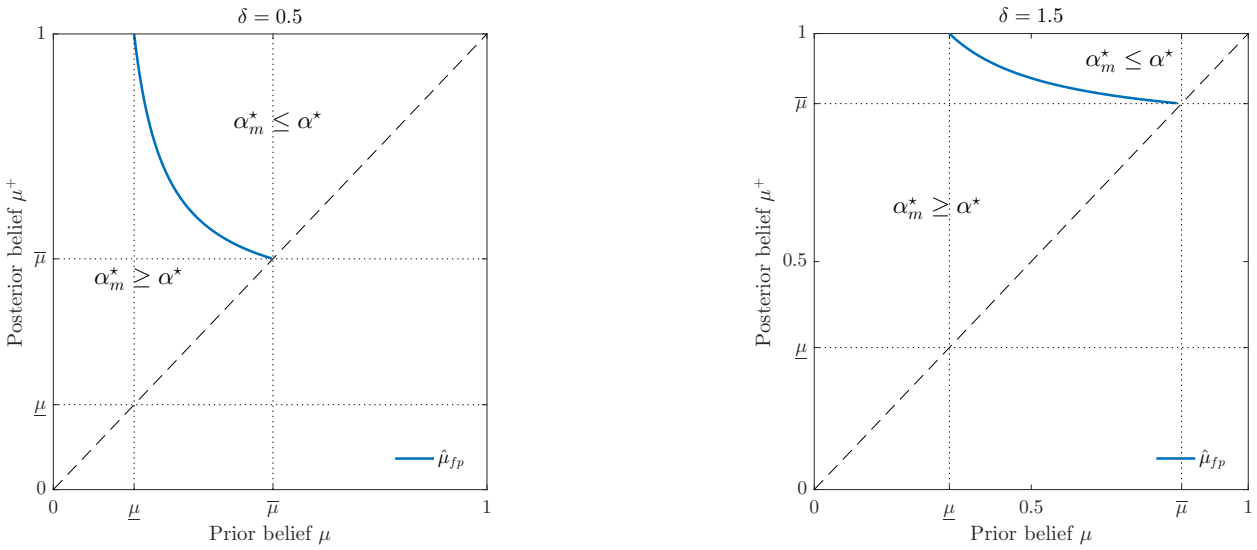


Figure 11 Impact of incentive structures on DM's false positive error rate ($\lambda = 1$)

Changing incentives also has a significant effect on the amount of cognitive cost that the DM incurs. In particular, the more at stake, the more cognitive effort the DM tolerates expending. As in our baseline case, the machine can only increase the DM's ex-ante effort when both her prior and posterior with the machine-supplied information induce the DM to exert cognitive effort (i.e., $\underline{\mu}, \mu < \mu^+, \overline{\mu}$). Therefore, as δ increases, the parameter region where the DM induces the DM to exert more cognitive effort increases as the difference $\overline{\mu} - \underline{\mu}$ becomes larger. This is illustrated in Figure 12.

Appendix C: Invariance of Accuracy to Prior Belief

We show this property by writing the DM's decision accuracy in terms of the optimal posteriors the DM constructs. The following lemma gives the characterization of these posteriors in the general payoff case.

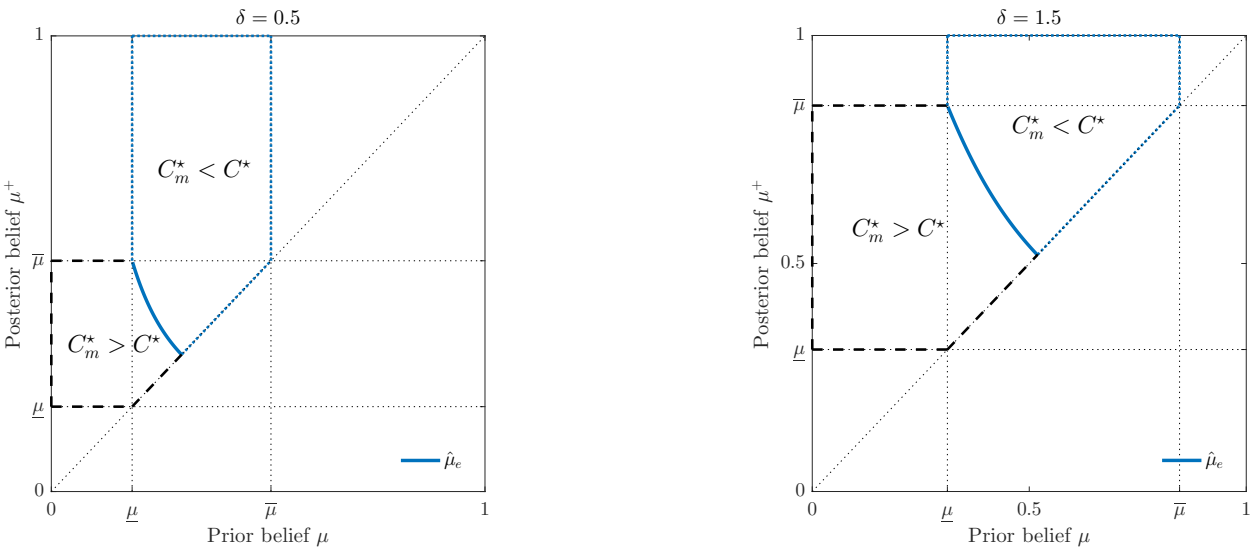


Figure 12 Impact of incentive structures on DM's cognitive effort ($\lambda = 1$)

LEMMA 3. *DMs optimal posterior beliefs when $\mu \in (\underline{\mu}, \overline{\mu})$ are*

$$\gamma(g|n) = \frac{1 - e^{-\delta/\lambda}}{e^{1/\lambda} - e^{-\delta/\lambda}}$$
$$\gamma(g|y) = \frac{1 - e^{-\delta/\lambda}}{e^{1/\lambda} - e^{-\delta/\lambda}} e^{1/\lambda}$$

with $\gamma(b|y) = 1 - \gamma(g|y)$ and $\gamma(b|n) = 1 - \gamma(g|n)$.

Proof We first find the optimal probability $p^\star$ of choosing $a = y$ for the general payoff case. By Theorem 1 in Matějka and McKay (2015), the DM's conditional probability of selecting $a = y$ given $\omega = g$ and $\omega = b$ are respectively, $P_g = \frac{pe^{1/\lambda}}{pe^{1/\lambda}+1-p}$ and $P_b = \frac{pe^{a/\lambda}}{pe^{a/\lambda}+(1-p)e^{c/\lambda}} = \frac{p}{p+(1-p)e^{\delta/\lambda}}$. Her unconditional choice probability p is then

$$p = (1-\mu)\frac{p}{p+(1-p)e^{\delta/\lambda}} + \mu\frac{pe^{1/\lambda}}{pe^{1/\lambda}+1-p}. \tag{25}$$

Solving (25) yields $\overline{p} = \frac{\mu}{1-e^{-\frac{1}{\lambda}\delta}} - \frac{1-\mu}{e^{\frac{1}{\lambda}}-1}$. Then, similar to the baseline model, $p^\star \le 0 \Leftrightarrow \mu \le \underline{\mu} = \frac{1-e^{-\frac{\delta}{\lambda}}}{e^{\frac{1}{\lambda}}-e^{-\frac{\delta}{\lambda}}}$ and $p^\star \ge 0 \Leftrightarrow \mu \ge \overline{\mu} = \frac{e^{\frac{1}{\lambda}}\left(1-e^{-\frac{\delta}{\lambda}}\right)}{e^{\frac{1}{\lambda}}-e^{-\frac{\delta}{\lambda}}}$. When $\mu \in [\underline{\mu}, \overline{\mu}]$, $p^\star = \overline{p}$. Using Bayes' rule, we have $\gamma(g|y) = p_g\mu/p^\star$ for the posterior belief that the state is good given $a = y$. Plugging in $p^\star$ and p_g, we arrive at $\gamma(g|y)$. Further we have $\gamma(b|y) = 1 - \gamma(g|y)$. The others are found similarly; $\gamma(g|n) = (1-p_g)\mu/(1-p^\star)$ and $\gamma(b|n) = 1 - \gamma(g|n)$. Q.E.D.

Writing the decision accuracy in terms of optimal posteriors $A(\mu) = \gamma(b|n)(1-p^\star) + \gamma(g|y)p^\star$, we see that when $\gamma(b|n) = \gamma(g|y)$, decision accuracy $A(\mu)$ does not depend on prior belief μ. Otherwise, it depends on μ through $p^\star$. By Lemma 3, $\gamma(b|n) = \gamma(g|y)$ if only if

$$\frac{e^{1/\lambda}-1}{e^{1/\lambda}-e^{-\delta/\lambda}} = \frac{1-e^{-\delta/\lambda}}{e^{1/\lambda}-e^{-\delta/\lambda}}e^{1/\lambda} \Leftrightarrow e^{-(\delta-1)/\lambda} = 1$$

which is only possible when $\delta = 1$. Here note that 1 refers to $u(y,g) - u(n,g)$, which is the gain from making the right decision in good state. Therefore when payoff gains across states are equal (i.e., symmetric), accuracy does not depend on prior belief μ.

Recent ESMT Working Papers

	ESMT No.
Contracting, pricing, and data collection under the AI flywheel effect Huseyin Gurkan, ESMT Berlin Francis de Véricourt, ESMT Berlin	20-01 (R1)
Queueing systems with rationally inattentive customers Caner Canyakmaz, ESMT Berlin Tamer Boyaci, ESMT Berlin	18-04 (R1)
The impact of EU cartel policy reforms on the timing of settlements in private follow-on damages disputes: An empirical assessment of cases from 2001 to 2015 Hans W. Friederiszick, ESMT Berlin and E.CA Economics Linda Gratz, E.CA Economics Michael Rauber, E.CA Economics	19-03 (R1)
Beyond retail stores: Managing product proliferation along the supply chain Işık Biçer, Schulich School of Business, York University Florian Lücker, Cass Business School, City, University of London Tamer Boyaci, ESMT Berlin	19-02 (R1)
Marginality, dividends, and the value in games with externalities Frank Huettner, ESMT Berlin André Casajus, HHL Leipzig Graduate School of Management	19-01