Boyacı, Tamer; Canyakmaz, Caner; de Véricourt, Francis

Working Paper

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

ESMT Working Paper, No. 20-02 (R1)

Provided in Cooperation with:
ESMT European School of Management and Technology, Berlin

Suggested Citation: Boyacı, Tamer; Canyakmaz, Caner; de Véricourt, Francis (2022) : Human and machine: The impact of machine input on decision-making under cognitive limitations, ESMT Working Paper, No. 20-02 (R1), European School of Management and Technology (ESMT), Berlin

This Version is available at:
https://hdl.handle.net/10419/266637

ISSN 1866-3494

November 29, 2022

ESMT Working Paper 20-02 (R1)

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

Tamer Boyaci, ESMT Berlin

Caner Canyakmaz, Ozyegin University

Francis de Véricourt, ESMT Berlin

Revised version

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

Tamer Boyacı
ESMT Berlin, 10178 Berlin, Germany, tamer.boyaci@esmt.org

Caner Canyakmaz
Faculty of Business, Ozyegin University, Istanbul 34794, Turkey, caner.canyakmaz@ozyegin.edu.tr

Francis de Véricourt
ESMT Berlin, 10178 Berlin, Germany, 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. Our main insights hold for different information and reward structures, and when the DM mistrust the machine.

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), human resources management (Gee 2017), supplier management (Saenz et al. 2020) or supply chain operations (IBM 2017).

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 2020). 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.

[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.

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 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, 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. However, 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. Yet, the machine cannot assess all information sources such as the DM's domain knowledge and the decision context.

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 (Mackowiak et al. 2021).

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 superiority of the human-machine collaboration, i.e., we show that the 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 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 different payoff structures, when the DM is biased against (or mistrusts) the machine, and when the machine is also imprecise. We explain in detail where and why they occur.

Our findings are most relevant in settings in which a human decision maker needs to exert some cognitive effort to make repetitive decisions that hinge on predictions. Examples include diagnostic tasks by radiologists (Liu et al. 2019), predictive maintenance and quality control in manufacturing (Brosset et al. 2019), the assessment of whether a part can be remanufactured in a production system (Nwankpa et al. 2021), evaluating applications by HR professionals (Gee 2017) or assessing a legal case in judicial systems (Cowgill 2018). Our framework is less suited, however, for decision tasks where the key unknown is a causal relationship.

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 ML 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 inn 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., 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. (2021), 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. Sun et al. (2021) also proposes a new bin packing algorithm that accounts for the tendency of human workers to deviate from machine's recommendations. Karlinsky-Shichor and Netzer (2019) find that providing an algorithm-based price recommendation to salespeople improves their pricing performances, which rely on their expertise, relationships and salesmanship skills. Conversely, Kesavan and Kushwaha (2020) find in a spare-parts retailer setting that allowing managers to deviate from the suggestion of an algorithm increases profitability. Others have further explored the conditions under which product category managers (Van Donselaar et al. 2010) or radiologists (Lebovitz et al. 2020) deviate from an algorithm recommendation.

Overall, these different streams of research focus on empircialy identifying when humans deviate from an algorithm's recommendation, and improving the interaction between humans and machinesA 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, i.e., 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 2021) among many others. Several empirical studies have further added support to the theory (see, for instance, Mackowiak et al. 2021 for a recent survey). Notably, Abis (2020) 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.

Perhaps closer to our paper, Jerath and Ren (2021) show in a standard rational inattention set-up that DMs always process information that confirms their prior beliefs when information cost is high. The machine's prediction in our setting interacts with this tendency to confirm a prior belief, leading sometimes the DM to actually exert more efforts when the information cost is high.

Besides rational inattention, other models of attention have been proposed. For instance Che and Mierendorff (2019) investigate a sequential attention allocation problem between two Poisson signals about a true state. However, the DM's information sources are restricted to these two signals in their model, while the DM has full flexibility to elicit any signal in our setup. In addition, the DM's information acquisition strategy is only driven by the incentive structure in Che and Mierendorff (2019), while this strategy is also determined by the DM's prior belief in our setting.

Our work is also related to the hypothesis testing Bayesian framework, in which the DM runs a series of imperfect tests and dynamically updates her belief accordingly about which decision is best (DeGroot 1970). This approach has been successfully applied to a variety of problems, such as the management of research projects or diagnostic services (McCardle et al. 2018, Alizamir et al. 2013, 2019), but is less suited to represent the cognitive process of a decision-maker. Indeed, this Bayesian framework typically assumes that each test's precision or the order in which they are run are exogenously determined. In contrast, our set-up fully endogenizes the level of precision as well as the associated cognitive effort in a tractable way. This enables to properly account for the flexibility of human cognition (Diamond 2013, Laureiro-Martínez and Brusoni 2018).

We further contribute to the nascent behavioral research on machine-human interactions, and the issue of trust in particular. For instance, Dietvorst et al. (2016) find in controlled experiments that DMs are adverse to machine-based predictions. de Véricourt and Gurkan (2022) also explore to which extent a DM may doubt a machine as she observes the correctness of its prescriptions

overtime. More generally, Donohue et al. (2020) call for more research in this field to better understand when human-machine collaborations provide superior preformance. We contribute to this by identifying environments, in which using a machine to improve efficiency is counter-productive.

Finally in our setup, humans assess information from multiple sources, which jointly designate the true state of the world. In this regard, our paper is related to the rich literature on search with multiple attributes (see, for instance, Olszewski and Wolinsky 2016, Sanjurjo 2017, and references therein). In particular, Huettner et al. (2019) study a multi-attribute discrete choice problem which generalizes the rational inattention model in Matějka (2015) to account for heterogenous 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.

[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 §6.1.

3.1. The Human Decision-Maker

The DM is constrained by her limited cognitive capacity, so that assessing available 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 prior μ to posterior $f(g|\mathbf{s})$. We measure uncertainty in terms of entropy, denoted as $H(p)$ for a probability p that the world is in the good state, where $H(p) = -p \log p - (1-p) \log (1-p)$. Entropy is a measure of uncertainty which corresponds to the expected 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

[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.

specifies the expected amount of elicited information. 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

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

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

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 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 perfectly assess the true state of the world.

The linearity of the cognitive cost in the expected reduction in entropy is a standard assumption in the rational inattention literature, which is justified by the fundamental coding theorem of information theory (see, e.g., Matějka and McKay 2015 for more details). Importantly, however, the cognitive cost is convex in the precision of the generated signals (i.e. C is convex in $f(\mathbf{s}|\omega)$). That is, eliciting additional more informative signals becomes increasingly costly.

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

[4] 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.

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:

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

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.

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. $-$) is indicative of a good (resp. bad) state. The true state of the world is good only if all available information is positive,[5] i.e.,

[5] 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.

$\omega = g$ 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). Thus, the DM's prior belief in the good state is $\mu = \pi(+, +)$. Without 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.

We consider a perfectly accurate machine for clarity only. As we discuss at the end of Section 6.2 and in Appendix D.1, our approach can be extended to account for inaccurate machine predictions. (All proofs and appendices can be found in the paper's electronic companion.)

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.[6] 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 $p \in [0, 1]$,*

$$p = (1 - \mu)\, p_b + \mu p_g, \quad \textit{where} \quad p_g = \frac{p e^{1/\lambda}}{p e^{1/\lambda} + 1 - p}, \quad p_b = \frac{p}{p + (1 - p)\, e^{1/\lambda}}. \tag{4}$$

[6] 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.

Further, we have

$$A^\star(\mu) = (1-\mu)(1-p_b) + \mu p_g \quad \textit{and} \quad C^\star(\mu) = \lambda\left[H(p) - (1-\mu)H(p_b) - \mu H(p_g)\right] \tag{5}$$

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 such that $\alpha^\star + \beta^\star = 1 - A^\star$, are equal to

$$\alpha^\star = (1-\mu)p_b \;\text{ and }\; \beta^\star = \mu(1-p_g). \tag{6}$$

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)*, is equal to*

$$p^\star(\mu) = \begin{cases} 0 & \textit{if } \mu \le \underline{\mu} \\ \frac{\mu}{1-e^{-1/\lambda}} - \frac{1-\mu}{e^{1/\lambda}-1} & \textit{if } \underline{\mu} < \mu < \overline{\mu} \\ 1 & \textit{if } \mu \ge \overline{\mu} \end{cases} \tag{7}$$

where $\underline{\mu} = (e^{1/\lambda}+1)^{-1} < 1/2 < \overline{\mu} = e^{1/\lambda}(e^{1/\lambda}+1)^{-1}$. 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 the 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*

$$p^\star(\lambda)=\begin{cases}\frac{\mu}{1-e^{-1/\lambda}}-\frac{1-\mu}{e^{1/\lambda}-1} & \text{if } \lambda<\bar{\lambda}\\ 0 & \text{if } \lambda\geq\bar{\lambda} \text{ and } \mu<0.5\\ 1 & \text{if } \lambda\geq\bar{\lambda} \text{ and } \mu>0.5,\end{cases} \tag{8}$$

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

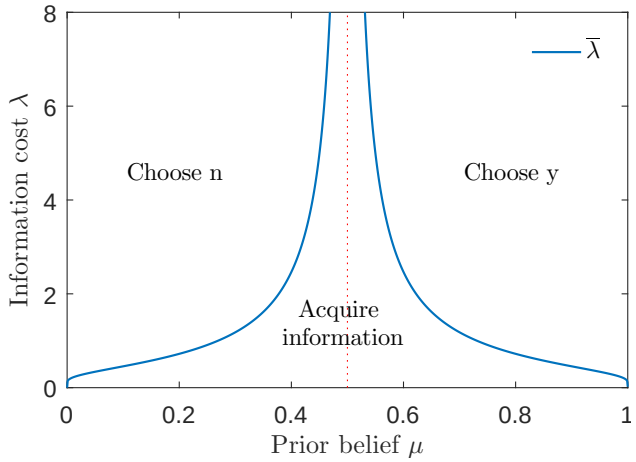


Figure 1 **Effect of prior belief μ on the DM's tolerance to information cost $\bar{\lambda}$**

Hence, the DM exerts effort only if the information cost is not too high, i.e., less than a threshold. In this case, her probability of choosing the good state increases with λ 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 the following closed forms for $A^\star(\mu)$, $C^\star(\mu)$ and $V^\star(\mu)$.

COROLLARY 2. *Given prior μ, we have*

- *If $\mu \le \underline{\mu}$, then $A^\star(\mu) = 1-\mu$ and $C^\star(\mu) = 0$.*
- *If $\underline{\mu} < \mu < \overline{\mu}$, then $A^\star(\mu) = \frac{e^{\frac{1}{\lambda}}}{e^{\frac{1}{\lambda}}+1}$ and $C^\star(\mu) = \lambda\left[H\left(\mu\right) - \varphi\left(\lambda\right)\right]$.*
- *If $\mu \ge \overline{\mu}$, then $A^\star(\mu) = \mu$ and $C^\star(\mu) = 0$, where*

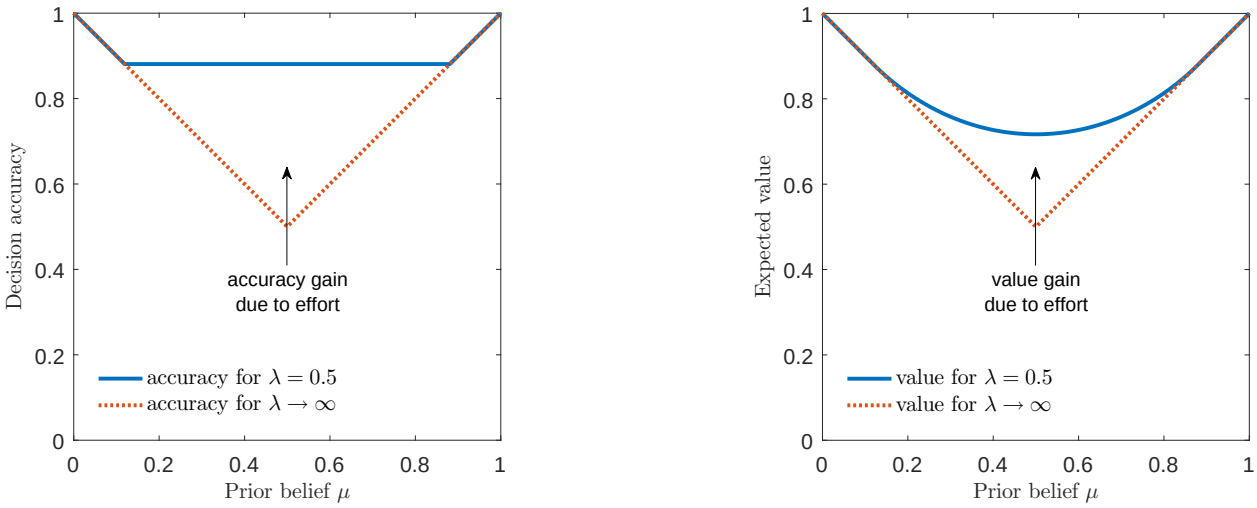
$$\varphi\left(\lambda\right) \equiv \log\left(e^{\frac{1}{\lambda}}+1\right) - \frac{1}{\lambda}\frac{e^{\frac{1}{\lambda}}}{e^{\frac{1}{\lambda}}+1}. \tag{9}$$

Further, $\varphi(\cdot)$ is increasing, with $\varphi(0) = 0$ and $\lim_{\lambda\to\infty}\varphi\left(\lambda\right) = \log 2$. Also, $V^\star(\mu) = A^\star(\mu) - C^\star(\mu)$.

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 independent of the prior. In fact, as long as the DM chooses to process information (i.e., $\underline{\mu} < \mu < \overline{\mu}$), her decision's expected accuracy 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 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 λ.

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 corresponds then to the expected gain that the DM enjoy for exerting cognitive effort.

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 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.



(a) Decision accuracy (b) Expected value

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

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). 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, 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 DM is bound to make choices based on partial information. Indeed, eliminating all uncertainty is never optimal ($\varphi(\lambda) > 0$ for $\lambda > 0$). Hence, accuracy is strictly less than one and the DM makes false positive and 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*

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

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, the DM processes some information and the error rates depend on both the prior and the information cost (with $0<\alpha^\star<1-\mu$ and $0<\beta^\star<\mu$).

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 increases, while the false negative one decreases as the prior increases. In fact, an increase in prior μ has two conflicting effects on the false positive rate. On one hand, the good state is more likely, which decreases the chance of false positive errors. 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 false positive errors. 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

With the 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 any given $\lambda > 0$ 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 showing 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 sometimes decrease with the machine (see Appendix F).

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 $\mu^+ \in (\underline{\mu}, \overline{\mu})$, 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 $\mu \in (\underline{\mu}, \overline{\mu})$ and $\mu^+ \geq \overline{\mu}$, then $\hat{\mu}_c$ exists such that $p_m^\star > p^\star$ if $\mu < \hat{\mu}_c$ 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 decreasing in μ^+ and equal to $\hat{\mu}_c = \left(e^{1/\lambda} + 1 - \frac{e^{1/\lambda}-1}{\mu^+}\right)^{-1} \geq 1/2$.

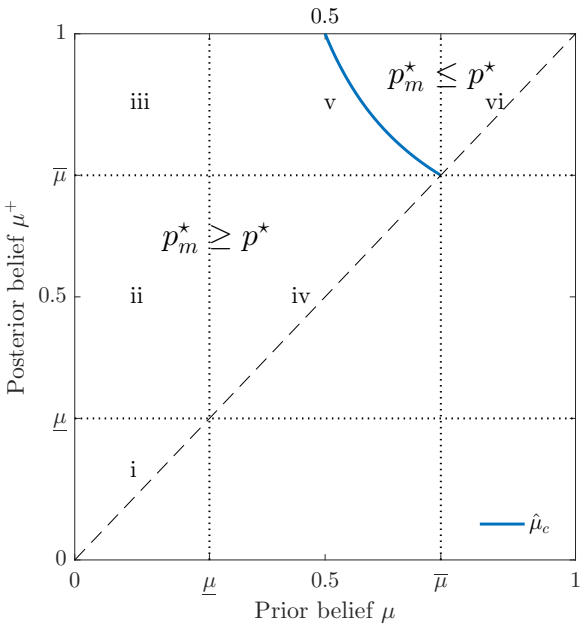


Figure 3 Impact of the machine on the DM's decision in parameter space S, for $\lambda = 1$.

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 ($\mu^+ \geq \bar{\mu}$). Threshold $\hat{\mu}_c$ is then the value of prior μ, at which the direction of the machine's impact changes.

Figure 3 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 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., $p_m^\star \leq p^\star$), 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* $p_m^\star \geq p^\star$.
- *If* $\mu > 0.5$, *then threshold* λ^* *exists such that* $p_m^\star \geq p^\star$ *if* $\lambda < \lambda^*$ *and* $p_m^\star \leq p^\star$ *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 4 illustrates the result and depicts threshold λ^* as a function of prior μ.

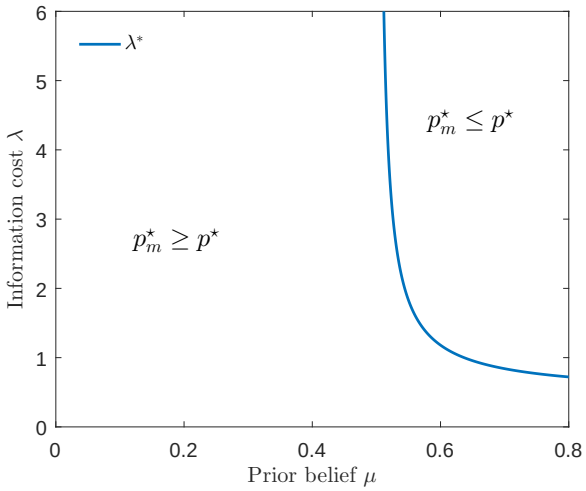


Figure 4 **Impact of the machine on the DM's decision as a function of information cost λ and prior μ, for $\mu^+ = 0.8$**

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 cost, the less information the DM assesses and thus the less likely it is that she will deviate from her prior choice.With 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 machine's assessment prompts the DM to immediately choose $a = y$ without exerting any additional effort (since $0.5 < \mu < \mu^+$). Thus, the ex-ante probability of choosing the good state, $p_m^\star$, 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.[7]

[7] 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.

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 increase 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.

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

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

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

iii) If $\mu \leq \underline{\mu}$ and $\mu^+ \geq \overline{\mu}$, 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 $\mu^+ \geq \overline{\mu}$, then threshold $\hat{\mu}_{fp} < \hat{\mu}_c$ exists such that $\alpha_m^\star > \alpha^\star$ if $\mu < \hat{\mu}_{fp}$, and $\alpha_m^\star \leq \alpha^\star$ otherwise.

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

Further, threshold $\hat{\mu}_{fp}$ is decreasing in μ^+ and equal to $\hat{\mu}_{fp} = \left(e^{2/\lambda} + e^{1/\lambda} - \frac{e^{2/\lambda}-1}{\mu^+}\right)^{-1}$.

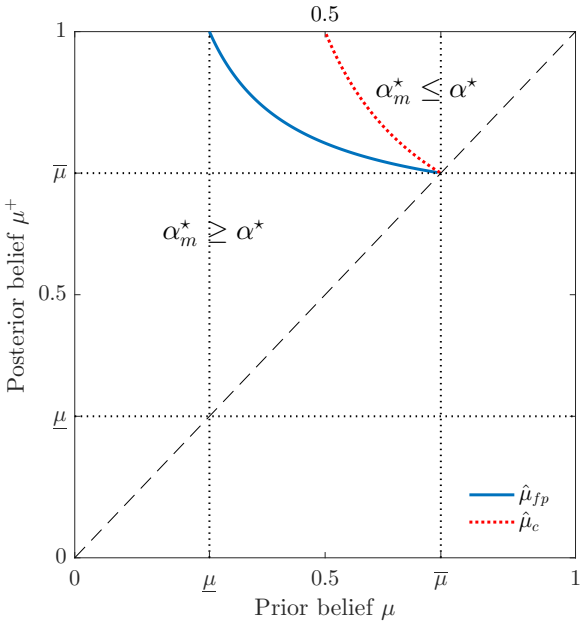


Figure 5 Impact of the machine on DM's false positive error rate in parameter space S, for $\lambda = 1$.

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., $p_m^\star \leq p^\star$ when $\mu \geq \mu_c^+$ 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 5 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, we $\alpha_m^\star \geq \alpha^\star$ for $\mu^+ \leq 0.5$. For $\mu^+ > 0.5$, we have*

- *If $\mu \leq \mu^* = 4\mu^+ \frac{1-\mu^+}{\left(2-\mu^+\right)^2}$, then $\alpha_m^\star \geq \alpha^\star$.*
- *If $\mu^* < \mu < 0.5$, $\underline{\lambda}_{fp}$ and $\overline{\lambda}_{fp}$ exist s.t. $\alpha_m^\star \geq \alpha^\star$ 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 $\alpha_m^\star \leq \alpha^\star$.*

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 not 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 an error. Yet, 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 6 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 Appendix A (proof of Corollary 5).

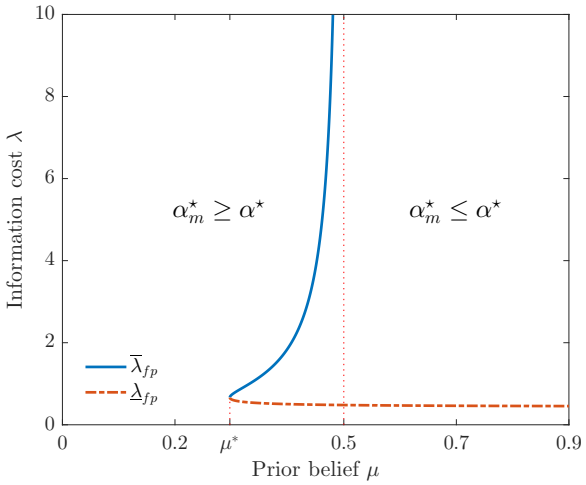


Figure 6 **Impact of the machine on DM's false positive error in information cost λ and prior μ, for $\mu^+ = 0.9$**

Taken together, the results of this section have important implications for the operations of organizations. In particular, the increase in false positives that the machine may induce translates into an unnecessary increase in capacity utilization in the downstream stages of a process. This means, for instance, that congestion levels and waiting times following the decision task can increase exponentially, as per basic queueing theory. This is indeed the case in manufacturing operations involving AI-assisted quality inspection and fault detection where increasing false positives may lead to unnecessary downtime, productivity losses and increased costs. Consequences are even more severe when the task consists in detecting low-frequency and high-risk events such as money laundering and fraud in banking, where even slight increases in false-positive rates may dramatically increase subsequent workload (Kaminsky and Schonert 2017).

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 $\mu^+ \le \underline{\mu}$, then $C_m^\star = C^\star = 0$.*

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

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

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

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

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

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

$$H(\mu) - \frac{\mu}{\mu^+} H(\mu^+) = (1 - \frac{\mu}{\mu^+})\varphi(\lambda) \tag{10}$$

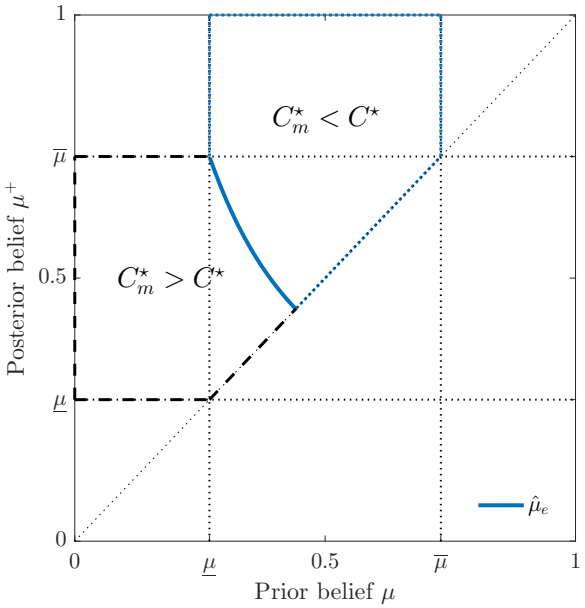


Figure 7 **Impact of the machine on DM's cognitive effort in parameter space $\mathcal{S}$, for $\lambda = 1$**

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 7. 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, which thus always reduces the task difficulty in expectation ($H(\mu) > \mathbb{E}_{X_1} H(\mu^{X_1})$). This task simplification contributes to reducing 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 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. Further, threshold λ_e^* is increasing in μ and satisfies*

$$\frac{H(\mu) - \frac{\mu}{\mu^+} H(\mu^+)}{1 - \frac{\mu}{\mu^+}} = \varphi(\lambda_e^*) \tag{11}$$

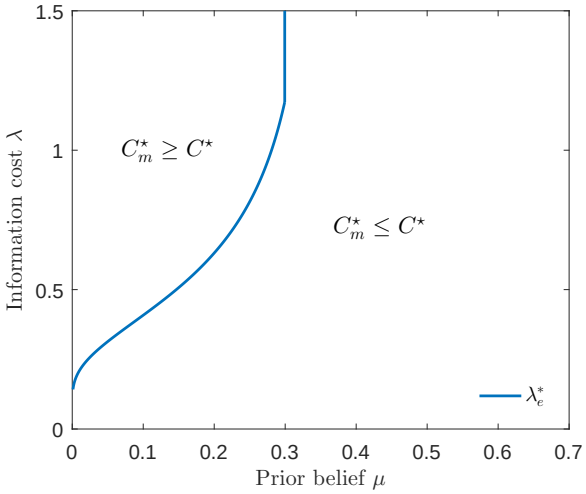


Figure 8 **Impact of the machine on DM's cognitive effort in information cost λ and prior μ, for $\mu^+ = 0.7$**

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 8 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. Extensions

We extend our baseline model in various directions to glean further insights regarding the impact of machine on human behavior. In particular, we first generalize the payoff structure and assume that DM aims to maximize her expected payoff (net of cognitive effort) instead of accuracy. This allows us to incorporate asymmetric costs for false negative and positive errors. We then explore settings where the DM mistrusts and is biased against the machine. Lastly, we study the case where the machine reduces the DM's uncertainty in a symmetric manner (i.e., not always revealing the bad state upon negative assessment), by considering three possible states of the world instead of two. More details and all formal results can be found in the Appendix.

6.1. 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 $(a,\omega) \in \{y,n\} \times \{g,b\}$. This general payoff structure may possibly create 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. We assume w.l.g that $u(y,g) = 1$ and $u(n,g) = 0$ (see Appendix B), such that difference $\delta = u(n,b) - u(y,b)$ represents the net value of correctly identifying the bad state.

We find that the threshold structure of our results continues to hold in this more general set-up (see Appendix B). Further, the set of beliefs μ and μ^+ for which $p_m^\star \geq p^\star$ widens as δ increases. Similar results hold for the false positive error rate and expected cognitive effort. 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. And as in our base case, the machine consistently reduces the false negative rate regardless of the incentive structure across states.

6.2. Incorporating Trust (Bias) to Machine Input

In our base model, the DM fully trusts the machine input. We extend our model to account for possible biases that the DM may hold against (or toward) the machine. As a result of this mistrust, the DM may not fully believe, for instance, that the state is bad when the machine's signal is negative. In this sense, the machine is not seen as perfectly accurate anymore. In fact, the framework we propose next can also account for the false positive or negative errors that an inaccurate machine may generate.

To account for the DM's trust and bias toward the machine, we follow the behavioral operations literature (see Özer et al. 2011) and assume that given machine input $x_1 \in \{+,-\}$, the DM updates her belief according to $\mu_\gamma^+ = (1-\gamma)\,\mu + \gamma\mu^+$ and $\mu_\gamma^- = (1-\gamma)\,\mu$, where higher values of trust parameter $\gamma \in [0,1]$ indicates more trust in the machine. That is, the DM mixes her prior belief μ with the posterior belief she would have were she to fully trust the machine. We retrieve our base model when $\gamma = 1$, while the DM fully ignores the machine and always decides alone when $\gamma = 0$. For $0 < \gamma < 1$, the DM's level of trust weakens the effect of the machine input on the DM's belief, i.e. $\mu^- = 0 < \mu_\gamma^- < \mu < \mu_\gamma^+ < \mu^+$. In particular, the negative signal of the machine does not fully reveal the bad state, that is $\mu_\gamma^- > 0$ in this case.

In this setup, we show that the machine always improves the DM's expected accuracy and value for any trust level (see Proposition 2, Appendix D). We also fully characterize the impact of the machine on the DM's behavior, and find that the machine may continue to increase the DM's propensity of making false positive errors. As in our base model, this happens when the DM does not strongly favor the good state a priori. In contrast to our base model, however, the machine may also increase the DM's propensity to make false negative errors. This happens when the DM strongly favors the good state a priori, and is due to the DM's mistrust in the machine's negative signal, which yields $\mu_\gamma^- > 0$ (see Proposition 3, Appendix D).

Similarly, we show that the machine can increase the DM's cognitive effort in this setup as well. This happens when the DM sufficiently favors either the bad or the good state (see Proposition 4 and Figure 13 in Appendix D). The former case is consistent with Theorem 4, and a similar rationale holds. The second case, however, does not occur in our base model.

6.3. A Symmetric Setting with an Additional State

In our base model, the machine reduces the DM's uncertainty in an asymmetric way as it fully resolves the bad state for the DM when the first information source X_1 is negative. We now extend our model to account for a symmetric setting, and show that our key insights continue to hold. In particular, we consider a third state, which we call *moderate* (denoted by $\omega = m$) and a corresponding accurate decision, which is declaring the test as *inconclusive* (denoted by $a = o$). We

assume that the true state is *good* (resp. *bad*) if and only if both X_1 and X_2 are positive (resp. negative). Otherwise (i.e., if $(X_1, X_2) \in \{(+,-),(-,+)\}$), the state is assumed to be *moderate*. In this setup, the machine never fully resolves the DM's uncertainty. That is, although a negative signal rules out the good state, the DM may still need to process information to distinguish the bad from the moderate state. Note that with more than two states, the DM now forms *consideration sets* (see Caplin et al. 2019) and may rule out some of them a priori, before eliciting any signal.

We fully characterize the DM's choice probabilities as a function of her prior beliefs in this set-up (Proposition 5 in Appendix E). As in our base model, we show that when information cost λ increases, the DM becomes less willing to process information for weaker prior beliefs (see Figure 14 in Appendix E). The machine also continues to always improves the DM's accuracy and expected value (Proposition 6, Appendix E).

Although the machine always improves overall accuracy, the machine may still increase certain error types and induce the DM to exert more cognitive effort as in our base model. To explore this, we focus on situations in which the machine reduces uncertainty in a symmetric manner, i.e. where the DM's prior and posterior beliefs are symmetric. In this case, we find that the machine increases the DM's false positive and negative errors, as well as her cognitive effort if she sufficiently favors the moderate state a priori (Propositions 7 and 8, Appendix E).[8]

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

[8] Accuracy increases because false positive and negative errors are offset by a decrease in false moderate errors.

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. Our results and insights are quite robust; they remain valid when the DM has a mistrust or bias against the machine assessment, and in generalized settings when the payoffs or machine impact on potential false positive and negative errors are altered.

Finally, we consider in this paper three different extensions of our base model, but others are possible. A noteworthy research direction is to explore how our findings change when the DM does not fully know the machine's accuracy. de Véricourt and Gurkan (2022) have recently proposed a dynamic bayesian framework to study this problem. A fruitful approach consists then in considering a setting similar to theirs, in which the DM is rationally inattentive as in ours.

Another interesting avenue of future research is the estimation and validation of our model using actual data. This could be conducted in a specific medical assessment setting, such as radiologists making diagnostic decisions with ML input from digital images. Another suitable setting is the sepsis alert system discussed in Ayvaci et al. (2021). Here, an algorithm (machine) studies the health status of a patient to generate an alert, which then triggers additional diagnostic actions by the caregivers to confirm sepsis detection. Different patient characteristics (e.g., age, disease history) naturally lead to different risk profiles regarding sepsis. These priors can be estimated

on the basis of past data. Through controlled experiments with and without machine input, it would be possible to study the changes in overall accuracy in detection, as well as the error rates. Conducting these experiments under varying time constraints, the impact of information costs can be determined. Combining such empirical results with the theoretical predictions would further advance our understanding of the conditions that make machine-based inputs most beneficial.

References

Abis S (2020) Man vs. machine: Quantitative and discretionary equity management. *SSRN 3717371* .

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* .

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.

Ayvaci MU, Mobini Z, Özer Ö (2021) To catch a killer: A data-driven personalized and compliance-aware sepsis alert system. *University of Texas at Dallas Working Paper* .

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.

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.

Brosset P, Jain A, Khemka Y, Buvat J, Thieullent A, Khadikar A, Patsko S (2019) Scaling ai in manufacturing operations: A practitioners' perspective. *Capgemini Research Institute* 10.

Canyakmaz C, Boyaci T (2021) Queueing systems with rationally inattentive customers. *Forthcoming in Manufacturing & Service Operations Management* .

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

Caplin A, Dean M, Leahy J (2019) Rational inattention, optimal consideration sets, and stochastic choice. *The Review of Economic Studies* 86(3):1061–1094.

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

Che YK, Mierendorff K (2019) Optimal dynamic allocation of attention. *American Economic Review* 109(8):2993–3029.

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.

de Véricourt F, Gurkan H (2022) Is your machine better than you? you may never know. Technical report, ESMT Berlin Working Paper 22-02.

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

DeGroot MH (1962) Uncertainty, information, and sequential experiments. *The Annals of Mathematical Statistics* 33(2):404–419.

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

Dietvorst BJ, Simmons JP, Massey C (2016) Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. *Management Science* 64(3):1155–1170, URL `http://dx.doi.org/10.1287/mnsc.2016.2643`.

Donohue K, Ozer O, Zheng Y (2020) Behavioral operations: Past, present and future. msom, 22(1) pp. 191-202, 2020. *Manufacturing & Service Operations Management* 11(22):191–202.

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.

Gee K (2017) In unilever's radical hiring experiment, resumes are out, algorithms are in. URL `https://www.wsj.com/articles/in-unilevers-radical-hiring-experiment-resumes-are-out-algorithms-are-in-1498478400`.

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.

IBM (2017) Welcome to the cognitive supply chain, executive report. URL `https://www.ibm.com/downloads/cas/DGP9YPZV`.

Ibrahim R, Kim SH, Tong J (2021) Eliciting human judgment for prediction algorithms. *Management Science* 67(4):2314–2325.

Jerath K, Ren Q (2021) Consumer rational (in) attention to favorable and unfavorable product information, and firm information design. *Journal of Marketing Research* 58(2):343–362.

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

Kaminsky P, Schonert J (2017) The neglected art of risk detection. Technical report.

Karlinsky-Shichor Y, Netzer O (2019) Automating the b2b salesperson pricing decisions: Can machines replace humans and when. *Available at SSRN* 3368402.

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.

Kesavan S, Kushwaha T (2020) Field experiment on the profit implications of merchants' discretionary power to override data-driven decision-making tools. *Management Science* 66(11):5182–5190.

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.

Lebovitz S, Lifshitz-Assaf H, Levina N (2020) To incorporate or not to incorporate ai for critical judgments: The importance of ambiguity in professionals' judgment process. *NYU Stern School of Business* .

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.

Mackowiak B, Matejka F, Wiederholt M (2021) Rational inattention: A review. *ECB Working Paper* .

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 (2018) When to abandon a research project and search for a new one. *Operations Research* 66(3):799–813.

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

Nwankpa C, Eze S, Ijomah W, Gachagan A, Marshall S (2021) Achieving remanufacturing inspection using deep learning. *Journal of Remanufacturing* 11(2):89–105.

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

Özer Ö, Zheng Y, Chen KY (2011) Trust in forecast information sharing. *Management Science* 57(6):1111–1137.

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, 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.

Saenz MJ, Revilla E, Simón C (2020) Designing ai systems with human-machine teams. *MIT Sloan Management Review* 61(3):1–5.

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`.

Sun J, Zhang DJ, Hu H, Van Mieghem JA (2021) Predicting human discretion to adjust algorithmic prescription: A large-scale field experiment in warehouse operations. *Management Science* .

Van Donselaar KH, Gaur V, Van Woensel T, Broekmeulen RA, Fransoo JC (2010) Ordering behavior in retail stores and implications for automated replenishment. *Management Science* 56(5):766–784.

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

Appendix A: Proofs of Results

This section contains proofs of all results in the main paper and the appendices.

Proof of Lemma 1 (4) follows 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), 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 (8) 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 (5) 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 (5) 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 $\mu \leq \underline{\mu}$, $p_b = p_g = 0$, hence $\alpha^\star = 0$ and $\beta^\star = \mu$ by (6). Similarly, when $\mu > \overline{\mu}$, $p_b = p_g = 1$, hence $\alpha^\star = 1-\mu$ and $\beta^\star = 0$. Now assume $\mu \in [\underline{\mu}, \overline{\mu}]$. Writing (6) in terms of optimal posteriors in Lemma 3 for $\delta = 1$ and plugging in optimal choice in (7), we obtain

$$\alpha^\star = (1-\mu)\, p_b = (1-\gamma(g|y))\, p = \frac{1}{e^{1/\lambda}+1}\left(\frac{\mu}{1-e^{-1/\lambda}} - \frac{1-\mu}{e^{1/\lambda}-1}\right) = \frac{\mu\left(e^{1/\lambda}+1\right)-1}{e^{2/\lambda}-1}$$

$$\beta^\star = \mu\,(1-p_g) = \gamma(g|n)\,(1-p) = \frac{1}{e^{1/\lambda}+1}\left(1 - \frac{\mu}{1-e^{-1/\lambda}} + \frac{1-\mu}{e^{1/\lambda}-1}\right) = \frac{e^{1/\lambda} - \mu\left(e^{1/\lambda}+1\right)}{e^{2/\lambda}-1}.$$

Proof of Lemma 2 Using $\mu^- = 0$, the result directly follows by (7), Corollary 2 and Corollary 3 for $p^\star$, $\alpha^\star$, $\beta^\star$, $A^\star$, $C^\star$, $V^\star$.

Proof of Proposition 1 $A^\star$ given in 2 is convex in μ for $[0,1]$. Then, by Jensen's inequality,

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

Similarly, the value function $V^\star$ is also convex in μ. To see this, note first that $V(\mu)$ is linearly decreasing in $[0,\underline{\mu}]$, convex in $[\underline{\mu},\overline{\mu}]$ (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 $V^\star(\mu) \leq V^\star_m(\mu)$.

Proof of Theorem 2 Note that *i*), *ii*), *iii*) and *vi*) follow by the optimal choice probability in (7) and the fact that $p^\star_m = \mu/\mu^+ p^\star(\mu^+)$.

iv) Using (7) and $\frac{\mu}{\mu^+} < 1$, we have

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

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

$$\frac{1}{e^{1/\lambda}-1} > \mu^+\left(\frac{1}{1-e^{-1/\lambda}} + \frac{1}{e^{1/\lambda}-1} - \frac{1}{\mu^+}\right). \tag{12}$$

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, (12) 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 $\mu \in [\underline{\mu}, \overline{\mu}]$ for a fixed λ. By Theorem 2, since $\widehat{\mu}_c \geq 0.5$, $p^\star_m \geq p^\star$ for $\mu \leq 0.5$. For $\mu < \underline{\mu}$, $p^\star_m \geq p^\star$ by *i*, *ii* and *iii*. This proves the first part. When $\mu > 0.5$, $p^\star_m \leq p^\star$ if $\mu \geq \widehat{\mu}_c$ for a fixed λ. Using $\hat{\mu}_c = \left(e^{1/\lambda} + 1 - \frac{e^{1/\lambda}-1}{\mu^+}\right)^{-1} \geq 1/2$ 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 Corollary 3 and the fact that $\alpha_m^\star = \mu/\mu^+ \alpha^\star(\mu^+)$ and $\beta_m^\star = \mu/\mu^{+\beta^\star}(\mu^+)$.

iv) Since $\frac{\mu}{\mu^+} < 1$ 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 Theorem 3 for λ, we obtain the following two roots;

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

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 Theorem 3 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 the DM's prior belief μ or posterior belief μ^+ induces her 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 $\mu^+\in\left(\widehat{\mu}_e^+,\overline{\mu}\right)$, the unique threshold $\widehat{\mu}_e$ satisfies

$$\frac{\widehat{\mu}_e}{\mu^+}\left(H\left(\mu^+\right)-\varphi\left(\lambda\right)\right)=H\left(\widehat{\mu}_e\right)-\varphi\left(\lambda\right). \tag{13}$$

Furthermore, left hand side of (13) is decreasing in μ^+ since

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

Therefore the crossing point that satisfies (13) and hence $\widehat{\mu}_e$ is decreasing in μ^+. Lastly, as the concave right hand side function in (13) 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 $H\left(\mu\right)>H\left(\mu^+\right)$ since H is symmetric around $\mu=0.5$. Then

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

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

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

Simplifying the numerator, we have

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

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, $-\left(1-\mu^+\right)\log\left(1-\mu^+\right)-\mu^+\log\mu^+-H\left(\mu^+\right)=0$. 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 (11).

Proof of Lemma 3 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

$$p=(1-\mu)\frac{p}{p+(1-p)e^{\delta/\lambda}}+\mu\frac{pe^{1/\lambda}}{pe^{1/\lambda}+1-p}. \tag{14}$$

Solving (14) 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, $p^\star\le 0\Leftrightarrow\mu\le\underline{\mu}$ and $p^\star\ge 1\Leftrightarrow\mu\ge\overline{\mu}$ where

$$\underline{\mu}=\frac{1-e^{-\frac{\delta}{\lambda}}}{e^{\frac{1}{\lambda}}-e^{-\frac{\delta}{\lambda}}}\quad\text{and}\quad\overline{\mu}=\frac{e^{\frac{1}{\lambda}}\left(1-e^{-\frac{\delta}{\lambda}}\right)}{e^{\frac{1}{\lambda}}-e^{-\frac{\delta}{\lambda}}}. \tag{15}$$

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.

Proof of Proposition 2 The accuracy function $A^\star(\mu)$ and value function $V^\star(\mu)$ given in Corollary 2 are convex in μ. Then by Jensen's inequality,

$$A^\star_m=P(X_1=+)A^\star(\mu^+_\gamma)+P(X_1=-)A^\star(\mu^-_\gamma)\ge A^\star(\mu)=A^\star$$

since $\mu=P(X_1=+)\mu^+_\gamma+P(X_1=-)\mu^-_\gamma$. The same holds for the expected value.

Proof of Theorem 5 Here we compare $p^\star=p^\star(\mu)$, $\alpha^\star=\alpha^\star(\mu)$, $\beta^\star=\beta^\star(\mu)$ and $C^\star=C^\star(\mu)$ with $p^\star_m=\frac{\mu}{\mu^+}p^\star\left(\mu^+_\gamma\right)+\left(1-\frac{\mu}{\mu^+}\right)p^\star\left(\mu^-_\gamma\right)$, $\alpha^\star_m=\frac{\mu}{\mu^+}\alpha^\star\left(\mu^+_\gamma\right)+\left(1-\frac{\mu}{\mu^+}\right)\alpha^\star\left(\mu^-_\gamma\right)$, $\beta^\star_m=\frac{\mu}{\mu^+}\beta^\star\left(\mu^+_\gamma\right)+\left(1-\frac{\mu}{\mu^+}\right)\beta^\star\left(\mu^-_\gamma\right)$ and $C^\star_m=\frac{\mu}{\mu^+}C^\star\left(\mu^+_\gamma\right)+\left(1-\frac{\mu}{\mu^+}\right)C^\star\left(\mu^-_\gamma\right)$ where $p^\star$ is given in Theorem 1, $C^\star$ in Corollary 2 and $\alpha^\star$ and $\beta^\star$ are given in Corollary 3. Note also that $\mu=\frac{\mu}{\mu^+}\mu^+_\gamma+\left(1-\frac{\mu}{\mu^+}\right)\mu^-_\gamma$.

i) (a), (b) and (c) follow since $p^\star(\mu)=\alpha^\star(\mu)=C^\star(\mu)=0$ for all $\mu\le\underline{\mu}$ and (c) follows since $\beta^\star$ is linear.

ii) (a), (b) and (d) follow since $p^\star(\mu)=\alpha^\star(\mu)=C^\star(\mu)=0$ and $p^\star\left(\mu_\gamma^+\right)=\alpha^\star\left(\mu_\gamma^+\right)=C^\star\left(\mu_\gamma^+\right)>0$. (c) follows by Jensen's inequality since $\beta^*(\mu)$ is concave in $[0,\overline{\mu}]$.

iii) (a) and (b) follow since $p^\star(\mu)=\alpha^\star(\mu)=0$ and $p^\star\left(\mu_\gamma^+\right)=\alpha^\star\left(\mu_\gamma^+\right)>0$. (d) follows since $C^\star(\mu)=C^\star\left(\mu_\gamma^+\right)=0$. (c) follows since $\beta^*\left(\mu_\gamma^+\right)=0$ and $\beta^*(\mu)>\beta^*\left(\mu_\gamma^-\right)$.

iv) (a), (b) and (c) follow since $p^*(\mu)$ and $\alpha^\star(\mu)$ are convex and $\beta^*(\mu)$ is concave in $[0,\overline{\mu}]$. (d) Let μ_γ^+ be fixed. Then, $C_m^\star=\frac{\mu}{\mu^+}C^\star\left(\mu_\gamma^+\right)$ is linearly increasing in μ from $\frac{\underline{\mu}}{\mu^+}C^\star\left(\mu_\gamma^+\right)$ to $C^\star\left(\mu_\gamma^+\right)$. Note also that $C^\star(\mu)$ is strictly concave in $\left[\underline{\mu},\mu_\gamma^+\right]$ starting from 0 to $C^\star\left(\mu_\gamma^+\right)$. This means two functions cross at a single point and $C_m^\star>C^\star$ when

$$C_m^\star=\frac{\mu}{\mu^+}C^\star\left(\mu_\gamma^+\right)>C^\star(\mu)\Longleftrightarrow\varphi(\lambda)>\frac{H(\mu)-\frac{\mu}{\mu^+}H\left(\mu_\gamma^+\right)}{1-\frac{\mu}{\mu^+}}.$$

v) (a), (b), (c) follow since $p^*(\mu)$, $\alpha^\star(\mu)$ and $\beta^*(\mu)$ are linear and $C^\star(\mu)$ is concave in $\left[\underline{\mu},\overline{\mu}\right]$.

vi) (a) $p_m^\star=\frac{\mu}{\mu^+}$ which is linearly increasing in μ. Similarly, $p^*(\mu)$ is linearly increasing in μ in $\left[\underline{\mu},\overline{\mu}\right]$. Note that slope of $\frac{\mu}{\mu^+}$ is lower than $p^*(\mu)$ in $\left[\underline{\mu},\overline{\mu}\right]$ since $p^*(\mu)$ goes from 0 to 1. Then they must cross at a single point. Furthermore, $p_m^\star>p^\star$ when

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

(b) Let μ_γ^+ be fixed. $\alpha^\star\left(\mu_\gamma^-\right)=0$ and $\alpha_m^\star=\frac{\mu}{\mu^+}(1-\mu_\gamma^+)$ is linearly increasing in μ. $\alpha^\star(\mu)$ is also linearly increasing in $\left[\underline{\mu},\overline{\mu}\right]$ with a larger slope (since $\alpha^\star\left(\underline{\mu}\right)=0$ and $\alpha^\star(\overline{\mu})$ is maximum). Then they cross at a single point. Furthermore, $\alpha_m^\star>\alpha^\star$ when

$$\alpha_m^\star=\frac{\mu}{\mu^+}(1-\mu_\gamma^+)>\frac{\mu\left(e^{1/\lambda}+1\right)-1}{e^{2/\lambda}-1}\Longleftrightarrow\mu<\frac{\frac{1}{e^{2/\lambda}-1}}{\frac{1}{e^{1/\lambda}-1}-\frac{1-\mu_\gamma^+}{\mu^+}}.$$

(c) Let μ_γ^+ be fixed. $\beta^\star(\overline{\mu})=0$ and $\beta_m^\star=\left(1-\frac{\mu}{\mu^+}\right)\beta^\star\left(\mu_\gamma^-\right)$ is linearly decreasing in μ.$\beta^\star(\mu)$ is also linearly decreasing in $\left[\underline{\mu},\overline{\mu}\right]$ with a larger slope (since $\beta^\star\left(\underline{\mu}\right)$ is maximum and $\beta^\star(\overline{\mu})=0$). Then they cross at a single point. Furthermore, $\beta_m^\star>\beta^\star$ when

$$\beta_m^\star=\left(1-\frac{\mu}{\mu^+}\right)\mu_\gamma^->\frac{e^{1/\lambda}-\mu\left(e^{1/\lambda}+1\right)}{e^{2/\lambda}-1}\Longleftrightarrow\mu>\frac{\frac{e^{1/\lambda}}{e^{2/\lambda}-1}-\mu_\gamma^-}{\frac{1}{e^{1/\lambda}-1}-\frac{\mu_\gamma^-}{\mu^+}}.$$

(d) follows since $C^\star(\mu)>0$ and $C^\star\left(\mu_\gamma^+\right)=C^\star\left(\mu_\gamma^-\right)=0$.

vii) (a), (b) and (c) follow since $p^*(\mu)$ and $\alpha^\star(\mu)$ are concave and $\beta^\star(\mu)$ is convex in $\left[\underline{\mu},1\right]$. (d) Let μ_γ^+ be fixed. Then, $C_m^\star=\left(1-\frac{\mu}{\mu^+}\right)C^\star\left(\mu_\gamma^-\right)$ is linearly decreasing in μ from $\frac{\underline{\mu}}{\mu^+}C^\star\left(\mu_\gamma^-\right)$ to $C^\star\left(\mu_\gamma^-\right)$. Note also that $C^\star(\mu)$ is strictly concave in $\left[\underline{\mu},\overline{\mu}\right]$ starting from $C^\star\left(\mu_\gamma^-\right)$ to 0. This means two functions cross at a single point and $C_m^\star>C^\star$ when

$$C_m^\star=\left(1-\frac{\mu}{\mu^+}\right)C^\star\left(\mu_\gamma^-\right)>C^\star(\mu)\Longleftrightarrow\frac{H(\mu)-\left(1-\frac{\mu}{\mu^+}\right)H\left(\mu_\gamma^-\right)}{\frac{\mu}{\mu^+}}<\varphi(\lambda).$$

viii) (a) follows since $p^\star(\mu) = p^\star(\overline{\mu}) = 1$ and $p^\star(\underline{\mu}) = 0$ (b) $\alpha^\star(\mu_\gamma^-) = 0$ and for $\mu \geq \overline{\mu}$, $\alpha^\star(\mu) > \alpha^\star(\mu_\gamma^+)$ as $\alpha^\star(\mu)$ is linearly decreasing in μ. Then $\alpha_m^\star = \frac{\mu}{\mu^+}\alpha^\star(\mu_\gamma^+) < \alpha^\star(\mu)$ (c) For $\mu \geq \overline{\mu}, \beta^\star(\mu_\gamma^+) = \beta^\star(\mu) = 0$. Since $\beta^\star(\mu_\gamma^-) > 0$, $\beta_m^\star > \beta^\star = 0$ (d) $C^\star(\mu) = 0$ and $C_m^\star$ since $C^\star(\mu_\gamma^+) = C^\star(\mu_\gamma^-) = 0$.

ix) (a), (b) and (c) follow since $p^*(\mu)$ and $\alpha^\star(\mu)$ are concave and $\beta^\star(\mu)$ is convex in $[\underline{\mu}, 1]$. (d) follows since $C^\star(\mu_\gamma^+) = C^\star(\mu) = 0$ and $C^\star(\mu_\gamma^-) > 0$.

x) (a) follows since $p^\star(\mu) = p^\star(\mu_\gamma^-) = p^\star(\mu_\gamma^+) = 1$. (b) follows since $\alpha^\star(\mu)$ is linear in μ. (c) follows since $\beta^\star(\mu) = \beta^\star(\mu_\gamma^-) = \beta^\star(\mu_\gamma^+) = 0$. (d) follows since $C^\star(\mu) = C^\star(\mu_\gamma^-) = C^\star(\mu_\gamma^+) = 0$.

Proof of Proposition 3 We use Theorem 5 in Appendix D. Note that $\alpha_m^\star > \alpha^\star$ in cases ii-iv. The union of the regions defined by ii and iii can be represented by $\{\mu < \underline{\mu}\}\&\{\mu_\gamma^+ > \underline{\mu}\}$, or equivalently $(\underline{\mu} - \gamma\mu^+)/(1-\gamma) < \mu < \underline{\mu}$ since $\mu_\gamma^+ = (1-\gamma)\mu + \gamma\mu^+$. Similarly, case iv and case vi for $\alpha_m^\star > \alpha^\star$ collectively imply the region $\{\mu_\gamma^- < \underline{\mu}\}\&\{\mu_\gamma^+ > \underline{\mu}\}\&\{\mu \in [\underline{\mu}, \overline{\mu}]\}\&\{\mu < \mu_{fp}^\gamma\}$, or equivalently, $\max\{\underline{\mu}, (\underline{\mu} - \gamma\mu^+)/(1-\gamma)\} < \mu < \min\{\underline{\mu}/(1-\gamma), \mu_{fp}^\gamma\}$. These two regions imply then that $\alpha_m^\star > \alpha^\star$ if $(\underline{\mu} - \gamma\mu^+)/(1-\gamma) < \mu < \widehat{\mu}_{fp}^\gamma$ where $\widehat{\mu}_{fp}^\gamma = \min\{\underline{\mu}/(1-\gamma), \mu_{fp}^\gamma\}$ where μ_{fp}^γ is given in Theorem 5. The same procedure applies for the false negatives.

Proof of Proposition 4 The conditions in cases ii and iv in Theorem 5 for $C_m^\star > C^\star$ collectively imply $(\underline{\mu} - \gamma\mu^+)/(1-\gamma) < \mu < \min\{(\overline{\mu} - \gamma\mu^+)/(1-\gamma), \underline{\mu}/(1-\gamma), \mu_e^l\}$. We define $\widehat{\mu}_e^{L\gamma} = \min\{(\overline{\mu} - \gamma\mu^+)/(1-\gamma), \underline{\mu}/(1-\gamma), \mu_e^l\}$. Similarly, cases vii and ix for $C_m^\star > C^\star$ imply $\max\{(\overline{\mu} - \gamma\mu^+)/(1-\gamma), \underline{\mu}/(1-\gamma), \mu_e^h\} < \mu < \overline{\mu}/(1-\gamma)$. We define $\widehat{\mu}^{H\gamma} = \max\{(\overline{\mu} - \gamma\mu^+)/(1-\gamma), \underline{\mu}/(1-\gamma), \mu_e^h\}$.

Proof of Proposition 5 The consumer problem 1 investigated in Caplin et al. (2019) can be used for our setup as well by taking $M = 3$, $u_G = 1$, $u_B = 0$, $\mu(\omega_k) = \mu_k$ and $\delta = e^{1/\lambda} - 1$. Then we apply Theorem 1 in Caplin et al. (2019) to arrive at our characterization.

Proof of Proposition 6 We first show that the accuracy function in (17) in Appendix E can be written as $A^\star(\mu_b, \mu_g) = \max\left\{\mu_1, \frac{e^{1/\lambda}(\mu_1+\mu_2)}{e^{1/\lambda}+1}, \frac{e^{1/\lambda}}{e^{1/\lambda}+2}\right\}$ in the domain $\mu_b + \mu_g \in [0,1]$. Assume $\mu_3 = 1 - \mu_1 - \mu_2 < \frac{1}{e^{1/\lambda}+2}$ and $\mu_1 > \mu_2 e^{1/\lambda}$. This is equivalent to $\mu_1 > \max\left\{\mu_2 e^{1/\lambda}, \frac{e^{1/\lambda}+1}{e^{1/\lambda}+2} - \mu_2\right\}$. Assume $\mu_2 > \frac{1}{e^{1/\lambda}+2}$. Then $\mu_1 > \mu_2 e^{1/\lambda} > \frac{e^{1/\lambda}+1}{e^{1/\lambda}+2} - \mu_2$ and since $\mu_2 e^{1/\lambda} > \frac{e^{1/\lambda}}{e^{1/\lambda}+2}$, we have $\mu_1 > \frac{e^{1/\lambda}}{e^{1/\lambda}+2}$. Assume now that Assume $\mu_2 \leq \frac{1}{e^{1/\lambda}+2}$. Then $\mu_1 > \frac{e^{1/\lambda}+1}{e^{1/\lambda}+2} - \mu_2 > \mu_2 e^{1/\lambda}$. Since $\frac{e^{1/\lambda}+1}{e^{1/\lambda}+2} - \mu_2 = \frac{e^{1/\lambda}}{e^{1/\lambda}+2} + \frac{1}{e^{1/\lambda}+2} - \mu_2 \geq \frac{e^{1/\lambda}}{e^{1/\lambda}+2}$ we also have $\mu_1 > \frac{e^{1/\lambda}}{e^{1/\lambda}+2}$. Note also that $\mu_1 > \mu_2 e^{1/\lambda}$ implies $\mu_1\left(e^{1/\lambda}+1\right) > e^{1/\lambda}(\mu_1+\mu_2)$, or equivalently $\mu_1 > \frac{e^{1/\lambda}(\mu_1+\mu_2)}{e^{1/\lambda}+1}$. Assume now that $\mu_3 > \frac{1}{e^{1/\lambda}+2}$. This implies that $1 - \mu_1 - \mu_2 > \frac{1}{e^{1/\lambda}+2} \iff \mu_1 + \mu_2 < \frac{e^{1/\lambda}+1}{e^{1/\lambda}+2} \iff \frac{e^{1/\lambda}}{e^{1/\lambda}+2} > \frac{e^{1/\lambda}(\mu_1+\mu_2)}{e^{1/\lambda}+1}$. When the conditions are reversed, the inequalities are also reversed. This proves that $A^\star(\mu_b, \mu_g) = \max\left\{\mu_1, \frac{e^{1/\lambda}(\mu_1+\mu_2)}{e^{1/\lambda}+1}, \frac{e^{1/\lambda}}{e^{1/\lambda}+2}\right\}$ in the domain $\mu_b + \mu_g \in [0,1]$. Then since $\mu_1, \frac{e^{1/\lambda}(\mu_1+\mu_2)}{e^{1/\lambda}+1}$ and $\frac{e^{1/\lambda}}{e^{1/\lambda}+2}$ are all convex functions, then $A^\star(\mu_b, \mu_g)$ is also convex since maximum function is preserved under convexity. As the accuracy function is convex, by Jensen's inequality the machine always increases the DM's expected decision accuracy as $(\mu_b, \mu_g) = P(X_1 = +)\left(0, \mu_g^+\right) + P(X_1 = -)\left(\mu_b^-, 0\right)$ and $A_m^\star = P(X_1 = +)A^\star\left(0, \mu_g^+\right) + P(X_1 = -)A^\star\left(\mu_b^-, 0\right) \geq A^\star(\mu_b, \mu_g)$. A similar approach is also valid for the value function $V^\star$.

Proof of Proposition 7 Assume $\mu_b = \mu_g = \mu$. We first use Proposition 5 to find the choice probabilities for the fully symmetric case. First assume that $\mu \leq \frac{1.}{e^{1/\lambda}+2}$. Then since $\mu < 1/3$, $\mu_1 = 1 - 2\mu$ and $\mu_2 = \mu_3 = \mu$. Also, since $1 - 2\mu > \mu e^{1/\lambda}$, by Proposition 5, $p_1^\star = p_o^\star = 1$ and $p_y^\star = p_n^\star = 0$. Now assume that $\frac{1.}{e^{1/\lambda}+2} < \mu \leq \frac{1}{3}$. Then, we have $p_y^\star = p_n^\star = \frac{\mu\left(e^{1/\lambda}+2\right)-1}{e^{1/\lambda}-1}$. Now assume $\mu > 1/3$. Then $\mu_1 = \mu_2 = \mu$ and $\mu_3 = 1 - 2\mu$. Note also

that $\mu_3 = 1-2\mu > \frac{1}{e^{1/\lambda}+2}$ implies $\mu < \frac{1}{2}\frac{e^{1/\lambda}+1}{e^{1/\lambda}+2}$. By Proposition 5, this means for $\frac{1}{3} < \mu < \frac{1}{2}\frac{e^{1/\lambda}+1}{e^{1/\lambda}+2}$, we have $p_y^\star = p_n^\star = \frac{\mu\left(e^{1/\lambda}+2\right)-1}{e^{1/\lambda}-1}$. Finally, if $\mu \geq \frac{1}{2}\frac{e^{1/\lambda}+1}{e^{1/\lambda}+2}$, we have $p_y^\star = p_n^\star = \frac{\mu\left(e^{1/\lambda}+2\right)-1}{e^{1/\lambda}-1}$ due to symmetry. Then note that $p_y^\star(\mu,\mu)$ the probability of choosing y can be written as a function of μ as

$$p_y^\star(\mu,\mu) = \begin{cases} 0 & \text{if } \mu \leq \frac{1}{e^{1/\lambda}+2} \\ \frac{\mu\left(e^{1/\lambda}+2\right)-1}{e^{1/\lambda}-1} & \text{if } \frac{1}{e^{1/\lambda}+2} < \mu < \frac{1}{2}\frac{e^{1/\lambda}+1}{e^{1/\lambda}+2} \\ \frac{1}{2} & \text{if } \mu \geq \frac{1}{2}\frac{e^{1/\lambda}+1}{e^{1/\lambda}+2} \end{cases} \tag{16}$$

The DM makes a false positive probability when she chooses y when the state is either bad or moderate. In optimality, false positive error probability is $\alpha^\star(\mu,\mu) = p_y^\star(\mu,\mu)\left(1-p_{g|y}^\star\right)$ where $p_{g|y}^\star$ is the posterior probability that the state is good given that the DM chooses y. From the optimal posteriors in Appendix E, we have $p_{g|y}^\star = \frac{e^{1/\lambda}}{e^{1/\lambda}+2}$ if all actions are chosen with positive probability (i.e., when $\frac{1}{e^{1/\lambda}+2} < \mu < \frac{1}{2}\frac{e^{1/\lambda}+1}{e^{1/\lambda}+2}$), $p_{g|y}^\star = \frac{2\mu e^{1/\lambda}}{e^{1/\lambda}+1}$ if only two options (y and n) are selected (i.e., when $\mu \geq \frac{1}{2}\frac{e^{1/\lambda}+1}{e^{1/\lambda}+2}$). Finally $p_{g|y}^\star = \mu$ if only one action is selected (i.e., only o is selected which happens when $\mu \leq \frac{1}{e^{1/\lambda}+2}$). Then the false positive rate as a function of μ becomes

$$\alpha^\star(\mu,\mu) = \begin{cases} 0 & \text{if } \mu \leq \frac{1}{e^{1/\lambda}+2} \\ 2\frac{\mu\left(e^{1/\lambda}+2\right)-1}{\left(e^{1/\lambda}-1\right)\left(e^{1/\lambda}+2\right)} & \text{if } \frac{1}{e^{1/\lambda}+2} < \mu < \frac{1}{2}\frac{e^{1/\lambda}+1}{e^{1/\lambda}+2} \\ \frac{1}{2} - \mu\frac{e^{1/\lambda}}{e^{1/\lambda}+1} & \text{if } \mu \geq \frac{1}{2}\frac{e^{1/\lambda}+1}{e^{1/\lambda}+2} \end{cases}.$$

When the machine provides $x_1 = +$ (resp. $x_1 = -$), then the DM has only two options good (resp. bad) and moderate with $\mu_g^+ = 2\mu$ and $\mu_m^+ = 1-2\mu$ (resp. $\mu_b^- = 2\mu$ and $\mu_m^- = 1-2\mu$). This means the DM can only make false positive errors when $x_1 = +$ which happens with probability 1/2. To find the false positive error rate in this case, we use the characterization in Corollary 3. Note that both error characterizations have two belief thresholds. Then we have the following different cases:

- If $\mu \leq \frac{1}{2}\frac{1}{e^{1/\lambda}+1}$, then $2\mu < \frac{1}{e^{1/\lambda}+1}$ and $\alpha_m^\star = \alpha^\star = 0$.
- If $\frac{1}{2}\frac{1}{e^{1/\lambda}+1} < \mu \leq \frac{1}{e^{1/\lambda}+2}$, then $2\mu > \frac{1}{e^{1/\lambda}+1}$ and hence $\alpha_m^\star > \alpha^\star = 0$. Note also that $\frac{1}{e^{1/\lambda}+2} > \frac{1}{2}\frac{1}{e^{1/\lambda}+1}$ for each $\lambda > 0$.
- If $\frac{1}{e^{1/\lambda}+2} < \mu \leq \frac{1}{2}\frac{e^{1/\lambda}}{e^{1/\lambda}+1}$ (or if $\frac{1}{2}\frac{e^{1/\lambda}}{e^{1/\lambda}+1} \leq \mu < \frac{1}{e^{1/\lambda}+2}$ depending on which one is bigger) then $\alpha_m^\star > \alpha^\star$ when $\frac{1}{2}\frac{2\mu\left(e^{1/\lambda}+1\right)-1}{e^{2/\lambda}-1} > 2\frac{\mu\left(e^{1/\lambda}+2\right)-1}{\left(e^{1/\lambda}-1\right)\left(e^{1/\lambda}+2\right)}$ which (after some mathematical manipulation) reduces to $\mu < \frac{\frac{3}{2}e^{1/\lambda}+1}{\left(e^{1/\lambda}+1\right)\left(e^{1/\lambda}+2\right)}$.However, this is only valid when $\lambda < 1/\ln 2$ since otherwise the threshold is greater than $\frac{1}{2}\frac{e^{1/\lambda}}{e^{1/\lambda}+1}$.
- If $\frac{1}{2}\frac{e^{1/\lambda}}{e^{1/\lambda}+1} < \mu < \frac{1}{2}\frac{e^{1/\lambda}+1}{e^{1/\lambda}+2}$, then, $\alpha_m^\star > \alpha^\star$ if $\frac{1}{2}(1-2\mu) > 2\frac{\mu\left(e^{1/\lambda}+2\right)-1}{\left(e^{1/\lambda}-1\right)\left(e^{1/\lambda}+2\right)}$which reduces to $\mu < \frac{\frac{1}{2}\left(e^{1/\lambda}-1\right)\left(e^{1/\lambda}+2\right)+2}{\left(e^{1/\lambda}+2\right)\left(e^{1/\lambda}+1\right)}$. The threshold is only valid when $\lambda > 1/\ln 2$, since otherwise it is less than $\frac{1}{2}\frac{e^{1/\lambda}}{e^{1/\lambda}+1}$.
- If $\mu \geq \frac{1}{2}\frac{e^{1/\lambda}+1}{e^{1/\lambda}+2}$, then $\alpha_m^\star < \alpha^\star$ since $\frac{1}{2}(1-2\mu) < \frac{1}{2} - \mu\frac{e^{1/\lambda}}{e^{1/\lambda}+1}$ for each $\lambda > 0$.

Taken together, $\alpha_m^\star > \alpha^\star$ only if $\frac{1}{2}\frac{1}{e^{1/\lambda}+1} < \mu < \min\left\{\frac{\frac{3}{2}e^{1/\lambda}+1}{\left(e^{1/\lambda}+1\right)\left(e^{1/\lambda}+2\right)}, \frac{\frac{1}{2}\left(e^{1/\lambda}-1\right)\left(e^{1/\lambda}+2\right)+2}{\left(e^{1/\lambda}+2\right)\left(e^{1/\lambda}+1\right)}\right\} = \mu_{fp}^*$. Note that the first component is increasing in λ while the second one is decreasing and it can be verified that they intersect at $\lambda = 1.4427$ which gives a value of 1/3. This means the threshold cannot be greater than 1/3 due to the minimum function.

Proof of Proposition 8 Using the characterization for the general case in Appendix E, the cognitive effort function for the symmetric case can be written as

$$C^{\star}(\mu,\mu)=\begin{cases}0 & \text{if } \mu\leq\frac{1}{e^{1/\lambda}+2}\\ \lambda\left[H(\mu,\mu)-\varphi^{(2)}(\lambda)\right] & \frac{1}{e^{1/\lambda}+2}<\mu<\frac{1}{2}\frac{e^{1/\lambda}+1}{e^{1/\lambda}+2}\\ 2\lambda\mu\left[\ln 2-\varphi(\lambda)\right] & \mu\geq\frac{1}{2}\frac{e^{1/\lambda}+1}{e^{1/\lambda}+2}\end{cases}$$

where $\varphi^{()}(\lambda)$ and $\varphi(\lambda)^{(2)}$ are given in (19) and (20) in Appendix E. Finally, $H(\mu,\mu)=-2\mu\ln(\mu)-(1-2\mu)\ln(1-2\mu)$. We now compare the DM's cognitive effort with and without the machine. Note that when the machine gives any information $x_1=+$ or $x_1=-$ (which happens with probability 1/2 in the fully symmetric case), the DM has moderate state and good or bad state to consider respectively, which is our base model. Since the DM's posterior beliefs are equal at 2μ for either case and since the effort function in our base case is symmetric, the DM's expected cognitive effort in the fully symmetric case is $C_m^{\star}=C^{\star}(2\mu)$ in our base case (see Corollary 2). Then we can use an approach similar to the Proof of Proposition 7 . We have the following cases:

- If $\mu\leq\frac{1}{2}\frac{1}{e^{1/\lambda}+1}$, then $C_m^{\star}=C^{\star}=0$ since the DM does not process information with or without the machine.
- If $\frac{1}{2}\frac{1}{e^{1/\lambda}+1}<\mu\leq\frac{1}{e^{1/\lambda}+2}$, then $C_m^{\star}>C^{\star}=0$ since the DM does not process information without the machine, but the machine induces her to process information.
- If $\frac{1}{e^{1/\lambda}+2}<\mu\leq\frac{1}{2}\frac{e^{1/\lambda}}{e^{1/\lambda}+1}$, then $C_m^{\star}>C^{\star}$ if

$$H(2\mu,0)-\varphi(\lambda)>H(\mu,\mu)-\varphi^{(3)}(\lambda)$$
$$-2\mu\ln(2\mu)-(1-2\mu)\ln(1-2\mu)-\varphi(\lambda)>-2\mu\ln(\mu)-(1-2\mu)\ln(1-2\mu)-\varphi^{(3)}(\lambda)$$
$$\mu<\frac{\varphi^{(3)}(\lambda)-\varphi(\lambda)}{2\ln 2}.$$

Depending on the level of λ, this threshold may be greater than the end point $\frac{1}{2}\frac{e^{1/\lambda}}{e^{1/\lambda}+1}$. In that case, $C_m^{\star}>C^{\star}$ in the whole region.

- If $\frac{1}{2}\frac{e^{1/\lambda}}{e^{1/\lambda}+1}<\mu<\frac{1}{2}\frac{e^{1/\lambda}+1}{e^{1/\lambda}+2}$, then $0=C_m^{\star}<C^{\star}$ since the DM without the machine does not process information as $2\mu>\frac{e^{1/\lambda}}{e^{1/\lambda}+1}$.
- If $\mu\geq\frac{1}{2}\frac{e^{1/\lambda}+1}{e^{1/\lambda}+2}$, then $0=C_m^{\star}<C^{\star}$ as well. All together, $C_m^{\star}>C^{\star}$ if and only if $\frac{1}{2}\frac{1}{e^{1/\lambda}+1}<\mu<\mu_e^{*}$ where $\mu_e^{*}=\min\left\{\frac{1}{2}\frac{e^{1/\lambda}}{e^{1/\lambda}+1},\frac{\varphi^{(3)}(\lambda)-\varphi(\lambda)}{2\ln 2}\right\}$. Finally, note that the second component in μ_e^{*} is increasing and in the limit equal to $\ln(3/2)/\ln(4)<1/3$.

Appendix B: General Payoff Structure

Our base model assumes that the DM's payoff corresponds to the overall accuracy of her decisions. Our framework can also account for a general payoff structure of the form $u(a,\omega)$, for $(a,\omega)\in\{y,n\}\times\{g,b\}$. More specifically, as we show below, we can normalize any payoff structure $u(a,\omega)$ such that $u(y,g)=1$ and $u(n,g)=0$ without loss of generality. 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.

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

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

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 the Machine for General Payoffs

Figure 9 depicts the impact of the machine on the DM's decision (analogous to Figure 3) 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 $\overline{\mu}$.

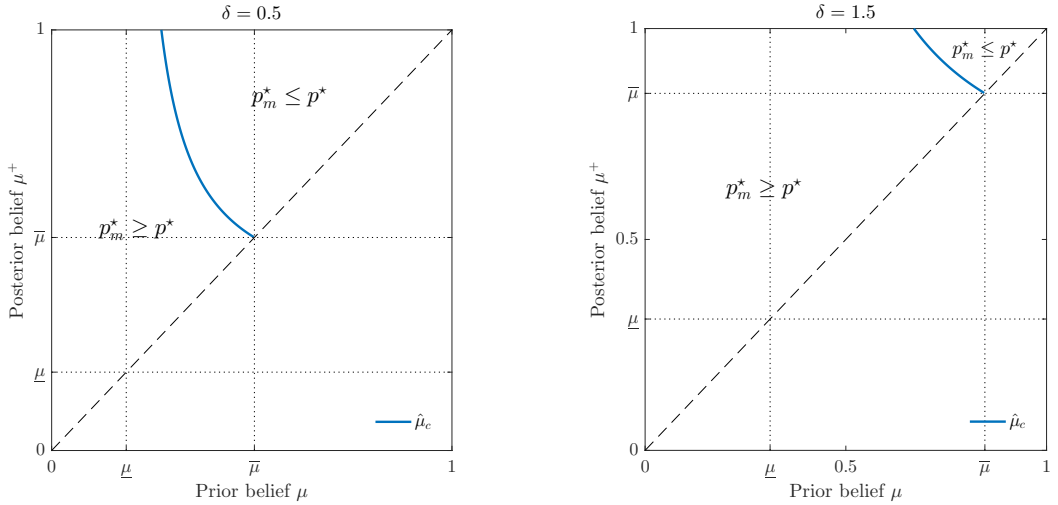


Figure 9 Impact of incentive structures on DM's decision ($\lambda=1$)

When DM's incentives change, the machine's impact on the extent of errors that the DM makes and the expected cognitive effort do 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 9). The effect of δ on DM's propensity to make false a positive error is illustrated in Figure 10.

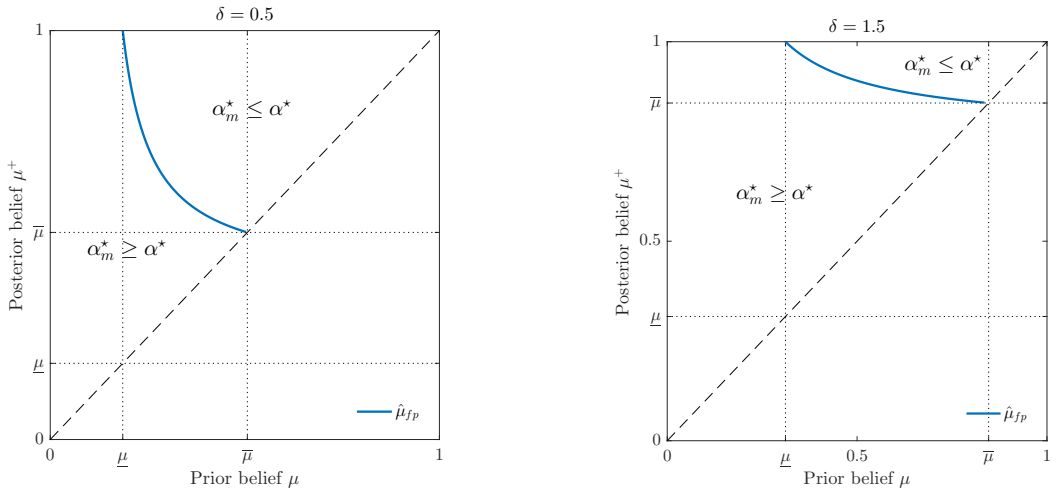


Figure 10 Impact of incentive structures on DM's false positive error rate ($\lambda = 1$)

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 11.

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.

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* $\gamma(b|n) = 1 - \gamma(g|n)$.

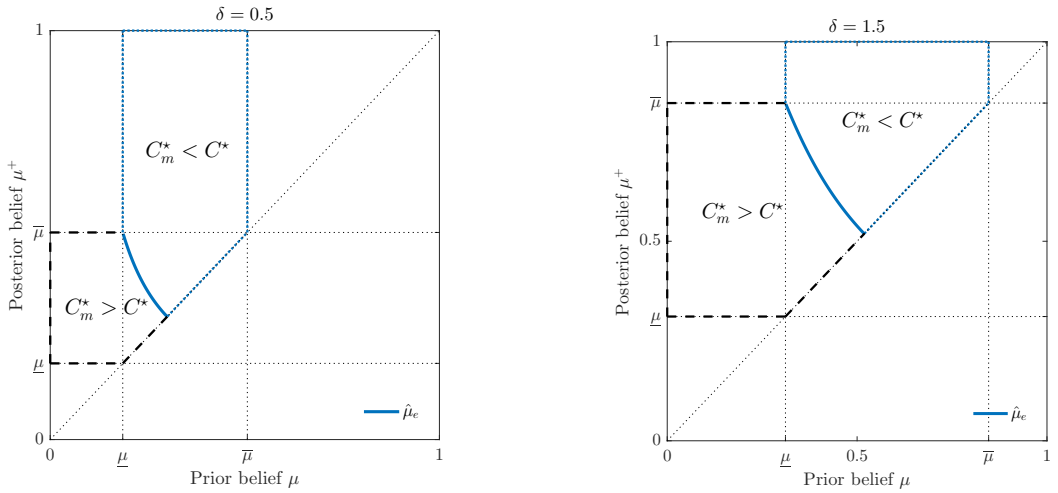


Figure 11 Impact of incentive structures on DM's cognitive effort ($\lambda = 1$)

Appendix D: Incorporating Trust

In this section we extend our baseline model to account for possible biases that the DM may hold against (or toward) the machine. The mistrust in machine implies that the DM may not fully believe that the state is bad when the machine's signal is negative. Hence, the machine is not seen as perfectly accurate. Indeed, as we explain at the end of section, our proposed framework also captures the case of an imprecise machine that may generate false positive or negative errors.

We follow the behavioral operations literature (e.g., Özer et al. 2011) in modeling DM's trust and bias towards the machine. Specifically, we assume that given machine input $x_1 \in \{+,-\}$, the DM updates her belief according to $\mu_\gamma^+ = (1-\gamma)\,\mu + \gamma\mu^+$ and $\mu_\gamma^- = (1-\gamma)\,\mu$, where higher values of trust parameter $\gamma \in [0,1]$ indicates more trust in the machine. In other words, the DM mixes her prior belief μ with the posterior belief she would have were she to fully trust the machine. Note that $\gamma = 1$ retrieves our base model, while the DM fully ignores the machine and always decides alone when $\gamma = 0$. For $0 < \gamma < 1$, the DM's level of trust weakens the effect of the machine input on the DM's belief, i.e. $\mu^- = 0 < \mu_\gamma^- < \mu < \mu_\gamma^+ < \mu^+$. In particular, the negative signal of the machine does not fully reveal the bad state, that is $\mu_\gamma^- > 0$ in this case. Nonetheless, we show next that the machine always improves the DM's expected accuracy and value for any trust level.

Proposition 2. *For any given $\lambda > 0$ and $\gamma \in [0,1]$, we have $A_m^\star \geq A^\star$ and $V_m^\star \geq V^\star$.*

To investigate the impact of the machine on the DM's behavior in this generalized setup, we follow the approach of Section 5 and first extend Theorem 1. This allows generalizing Theorems 2, 3, and 4 for any trust parameter $\gamma \in [0,1]$. Note that in addition to prior belief μ and posterior beliefs μ_γ^+, our results also depend on $\mu_\gamma^- > 0$ when $\gamma \in (0,1)$. Because of this, the number of parameter regions that describe the effect of the machine increases from six when $\mu^- = 0$, as in the base case, to ten when $\mu_\gamma^- > 0$. Theorem 5 provides a full characterization of the impact of the machine on the DM's choice probability, false positive/negative error rates and cognitive effort. Note that we assume $\mu_\gamma^- < \mu < \mu_\gamma^+$.

Theorem 5. *Given information cost $\lambda > 0$ and $\mu < \mu^+$, we have*

i) If $\mu_\gamma^+ \leq \underline{\mu}$*, then, (a)* $p_m^\star = p^\star = 0$*. (b)* $\alpha_m^\star = \alpha^\star = 0$*. (c)* $\beta_m^\star = \beta^\star > 0$*. (d)* $C_m^\star = C^\star = 0$*.*

ii) If $\mu \leq \underline{\mu}$ *and* $\mu_\gamma^+ \in \left(\underline{\mu}, \overline{\mu}\right)$*, then, (a)* $p_m^\star > p^\star = 0$*. (b)* $\alpha_m^\star > \alpha^\star = 0$*. (c)* $\beta_m^\star < \beta^\star$*. (d)* $C_m^\star > C^\star = 0$*.*

iii) If $\mu \leq \underline{\mu}$ *and* $\mu_\gamma^+ \geq \overline{\mu}$*, then (a)* $p_m^\star > p^\star = 0$*. (b)* $\alpha_m^\star > \alpha^\star = 0$*. (c)* $\beta_m^\star < \beta^\star$*. (d)* $C_m^\star = C^\star = 0$*.*

iv) If $\mu_\gamma^- \leq \underline{\mu}$ *and* $\mu, \mu_\gamma^+ \in \left[\underline{\mu}, \overline{\mu}\right]$*, then (a)* $p_m^\star > p^\star$*. (b)* $\alpha_m^\star > \alpha^\star$*. (c)* $\beta_m^\star < \beta^\star$*.*

(d) threshold $\widehat{\mu}_e^l$ *exists such that* $C_m^\star > C^\star$ *if* $\mu < \widehat{\mu}_e^l$ *and* $C_m^\star \leq C^\star$ *otherwise. Furthermore, threshold* $\widehat{\mu}_e^l$ *uniquely solves (for* μ*)*

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

v) If $\mu_\gamma^-, \mu, \mu_\gamma^+ \in \left[\underline{\mu}, \overline{\mu}\right]$*, then, (a)* $p_m^\star = p^\star$*. (b)* $\alpha_m^\star = \alpha^\star$*. (c)* $\beta_m^\star = \beta^\star$*. (d)* $C_m^\star < C^\star$

vi) If $\mu_\gamma^- \leq \underline{\mu}$ *and* $\mu \in \left[\underline{\mu}, \overline{\mu}\right]$ *and* $\mu_\gamma^+ \geq \overline{\mu}$*, then,*

(a) *threshold* $\widehat{\mu}_c$ *exists such that* $p_m^\star > p^\star$ *if* $\mu < \widehat{\mu}_c$ *and* $p_m^\star \leq p^\star$ *otherwise. Furthermore, threshold* $\widehat{\mu}_c$ *is given as*

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

(b) *threshold* $\widehat{\mu}_{fp}^\gamma$ *exists such that* $\alpha_m^\star > \alpha^\star$ *if* $\mu < \widehat{\mu}_{fp}^\gamma$ *and* $\alpha_m^\star \leq \alpha^\star$ *otherwise. Furthermore, the threshold* $\widehat{\mu}_{fp}^\gamma$ *uniquely solves (for* μ*)*

$$\frac{\mu\left(e^{1/\lambda} + 1\right) - 1}{e^{2/\lambda} - 1} = \frac{\mu}{\mu^+}\left(1 - \mu_\gamma^+\right)$$

(c) *threshold* $\widehat{\mu}_{fn}^\gamma$ *exists such that* $\beta_m^\star < \beta^\star$ *if* $\mu < \widehat{\mu}_{fn}^\gamma$ *and* $\beta_m^\star \geq \beta^\star$ *otherwise. Furthermore, the threshold* $\widehat{\mu}_{fn}^\gamma$ *uniquely solves (for* μ*)*

$$\frac{e^{1/\lambda} - \mu\left(e^{1/\lambda} + 1\right)}{e^{2/\lambda} - 1} = \left(1 - \frac{\mu}{\mu^+}\right)\mu_\gamma^-$$

(d) $0 = C_m^\star < C^\star$

vii) If $\mu_\gamma^-, \mu \in \left[\underline{\mu}, \overline{\mu}\right]$ *and* $\mu_\gamma^+ \geq \overline{\mu}$*, then, (a)* $p_m^\star < p^\star$*. (b)* $\alpha_m^\star < \alpha^\star$*. (c)* $\beta_m^\star > \beta^\star$*.*

(d) threshold $\widehat{\mu}_e^h$ *exists such that* $C_m^\star < C^\star$ *if* $\mu < \widehat{\mu}_e^h$ *and* $C_m^\star \geq C^\star$ *otherwise. Furthermore, threshold* $\widehat{\mu}_e^h$ *uniquely solves (for* μ*)*

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

viii) If $\mu_\gamma^- \leq \underline{\mu}$ *and* $\mu_\gamma^+ \mu \geq \overline{\mu}$*, then, (a)* $p_m^\star < p^\star = 1$*. (b)* $\alpha_m^\star < \alpha^\star$*. (c)* $\beta_m^\star > \beta^\star$*. (d)* $C_m^\star = C^\star = 0$*.*

ix) If $\mu_\gamma^- \in \left[\underline{\mu}, \overline{\mu}\right]$ *and* $\mu_\gamma^+, \mu \geq \overline{\mu}$*, then, (a)* $p_m^\star < p^\star = 1$*. (b)* $\alpha_m^\star < \alpha^\star$*. (c)* $\beta_m^\star > \beta^\star = 0$*. (d)* $C_m^\star > C^\star = 0$

x) If $\mu_\gamma^- \geq \overline{\mu}$*, then (a)* $p_m^\star = p^\star = 1$*. (b)* $\alpha_m^\star = \alpha^\star$*. (c)* $\beta_m^\star = \beta^\star = 0$*. (d)* $C_m^\star = C^\star = 0$*.*

The following results highlight how trust interacts with the effect of the machine on the DM's decision making process.

PROPOSITION 3. *For any given* $\lambda > 0$*,* $\gamma \in (0, 1]$ *and* $\mu^+ > \mu$*, thresholds* $\widehat{\mu}_{fp}^\gamma$ *and* $\widehat{\mu}_{fn}^\gamma$ *exist such that*

i) $\alpha_m^\star > \alpha^\star$ *if and only if* $\frac{\mu - \gamma\mu^+}{1-\gamma} < \mu < \widehat{\mu}_{fp}^\gamma$*,*

ii) $\beta_m^\star > \beta^\star$ *if and only if* $\widehat{\mu}_{fn}^\gamma < \mu < \frac{\overline{\mu}}{1-\gamma}$*.*

Proposition 3 shows that the machine may continue to increase the DM's propensity of making false positive errors in the presence of mistrust. As in our base model, this happens when the DM does not strongly favor the good state a priori (i.e., $\mu < \widehat{\mu}_{fp}^{\gamma}$). In this sense, the result in Theorem 3 is robust to the inclusion of trust. In contrast to our base model, however, the machine may also increase the DM's propensity to make false negative errors. This is due to the DM's mistrust in the machine's negative signal, which yields $\mu_{\gamma}^{-} > 0$. This happens when the DM strongly favors the good state a priori (i.e., $\mu > \widehat{\mu}_{fn}^{\gamma}$).

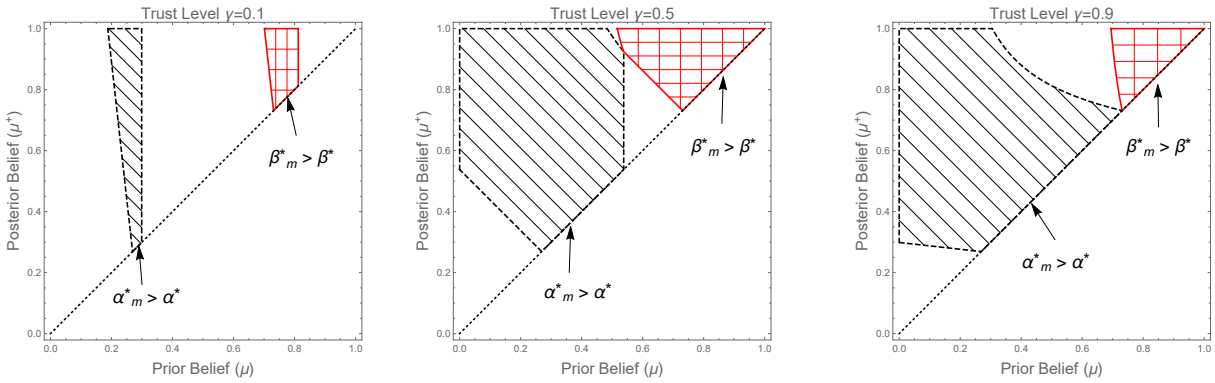


Figure 12 Impact of the machine on the DM's decision errors for different trust levels

Figure 12 illustrates the parameter regions where the machine increases the DM's errors for three different trust levels. Since the machine always increases DM's decision accuracy as per Proposition 2, the machine never increases both false positive and negative rates at the same time.

The DM's bias toward the machine also interacts with the effect the machine has on the DM's cognitive effort, as we show next.

PROPOSITION 4. *For any given $\lambda > 0$, $\gamma \in (0,1]$ and $\mu^{+} > \mu$, thresholds $\widehat{\mu}_{e}^{L\gamma}$ and $\widehat{\mu}_{e}^{H\gamma}$ exist such that $C_{m}^{\star} > C^{\star}$ if and only if $\frac{\mu-\gamma\mu^{+}}{1-\gamma} < \mu < \widehat{\mu}_{e}^{L\gamma}$ or $\widehat{\mu}_{e}^{H\gamma} < \mu < \frac{\overline{\mu}}{1-\gamma}$.*

Proposition 4 shows that the machine may increase the DM's cognitive effort in this setup as well. This can happen when the DM sufficiently favors either the bad state (i.e., $\mu < \widehat{\mu}_{e}^{L\gamma}$), or the good state (i.e., $\mu > \widehat{\mu}_{e}^{H\gamma}$). The former case is consistent with Theorem 4 in the base model, and a similar rationale holds. The second case, however, does not occur in our base model. This is because the DM's posterior belief upon a negative machine signal is positive if $\gamma > 0$. This may then correspond to an increase in task difficulty and thus induces the DM to process more information.

Figure 13 illustrates the parameter regions in which the machine increases the DM's cognitive effort for different values of γ. The figure depicts two such regions, one for each of the two intervals in μ defined by Proposition 4. Note that when trust level γ is sufficiently high, the region corresponding to the second interval disappears since posterior μ_{γ}^{-} approaches to zero, which corresponds to our base model.

A key point in the above analysis is that the DM's optimal choice probabilities, decision errors and cognitive efforts only depend on prior belief μ (without the machine) and posterior beliefs μ_{γ}^{+} and μ_{γ}^{-} (with the machine) in addition to λ. In our original setup $\mu_{\gamma}^{-} = 0$, but Theorem 5 extends these results to a setup

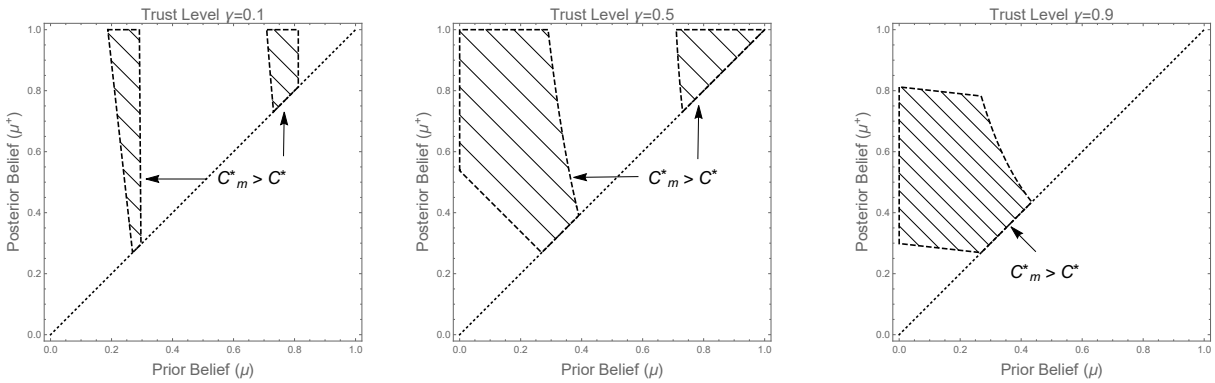


Figure 13 Impact of machine on DM's cognitive effort for different trust levels

where μ, μ_γ^+ and μ_γ^- are all free parameters, with $\mu_\gamma^- \geq 0$. Importantly, this characterization is context free, i.e., once these probabilities are specified, the previous approach holds. (Theorem 5 also provides thresholds $\widehat{\mu}_{fp}^{\gamma}$, $\widehat{\mu}_{fn}^{\gamma}$, $\widehat{\mu}_e^{L\gamma}$ and $\widehat{\mu}_e^{H\gamma}$ of Propositions 3 and 4 in closed form.).

D.1. Incorporating Machine Imprecision

Assume that the input the machine provides is not always accurate. In particular, assume that the machine gives signal $Y \in \{+,-\}$ with false positive error probability $\gamma_1 = P(Y=+|\omega=b)$ and false negative error probability $\gamma_2 = P(Y=-|\omega=g)$. Upon receiving a positive signal from the machine, the DM updates her prior as

$$\mu_\gamma^+ = P(\omega=g \,|\, Y=+) = \frac{(1-\gamma_2)\,\mu}{(1-\gamma_2)\,\mu+\gamma_1\,(1-\mu)}$$
$$\mu_\gamma^- = P(\omega=g \,|\, Y=-) = \frac{\gamma_2\mu}{\gamma_2\mu+(1-\gamma_1)\,(1-\mu)}.$$

Assuming $\gamma_1+\gamma_2<1$ ensures that $\mu_\gamma^- < \mu < \mu_\gamma^+$ and Theorem 5 can directly be used to assess the impact of the machine on the DM's choice behavior. If $\gamma_1+\gamma_2>1$, then $\mu_\gamma^+ < \mu < \mu_\gamma^-$ and it just enough to relabel μ_γ^+ and μ_γ^- and apply Theorem 5. Note that it is possible to combine mistrust and imprecision to model situations where the DM is biased *and* the machine is inaccurate.

Appendix E: A Symmetric Setting with an Additional State

In our base model, the machine reduces the DM's uncertainty in an asymmetric way as it fully resolves the bad state for the DM when the first information source X_1 is negative. In order to account for a symmetric reduction setting, we introduce a third state, which we call *moderate* (denoted by $\omega=m$) and a corresponding accurate decision, which is declaring the test as *inconclusive* (denoted by $a=o$). We assume that the true state is *good* (resp. *bad*) if and only if both X_1 and X_2 are positive (resp. negative). Otherwise (i.e., if $(X_1,X_2)\in\{(+,-),(-,+)\}$), the state is assumed to be *moderate*. In this setup, the machine never fully resolves the DM's uncertainty. That is, although a negative signal rules out the good state, the DM may still need to process information to distinguish the bad from the moderate state.

It is well-known (see Caplin et al. 2019) that with more than two states, the DM may naturally rule out some of the states a priori before eliciting any signal (i.e., forms *considerations sets*). To accommodate for this extension, we slightly change our notation and let $p_a^\star(\mu_b,\mu_g)$ denote the unconditional probability that

the DM chooses $a \in \{n, y, o\}$ as a function her prior belief $\mu_b = \pi(-,-)$ that the state is bad and $\mu_g = \pi(+,+)$ that the state is good. The DM's prior belief for the moderate state is then $\mu_m = 1 - \mu_b - \mu_g$.

We characterize next the DM's choice probabilities as a function of her prior belief for a given information cost λ. To that end, we order the priors such that $\mu_1 \equiv \max(\mu_b, \mu_g, \mu_o)$, $\mu_3 \equiv \min(\mu_b, \mu_g, \mu_o)$ and μ_2 denotes the remaining prior with $\mu_1 \geq \mu_2 \geq \mu_3$. Action $a_i, i \in \{1,2,3\}$ indicates then the action corresponding to the state associated with prior μ_i. For instance, if $\mu_1 = \mu_g$, $\mu_2 = \mu_b$ and $\mu_3 = \mu_m$ then $a_1 = y$, $a_2 = n$ and $a_3 = o$. We also have $\mu_m = 1 - \mu_b - \mu_g$ and thus present our result in the parameter space $\mu_g \times \mu_b$.

PROPOSITION 5. *Let* $\mu_3 = \min\{\mu_g, \mu_b, 1-\mu_b-\mu_g\}$, $\mu_1 = \max\{\mu_g, \mu_b, 1-\mu_b-\mu_g\}$, *and* $\mu_2 = 1-\mu_1-\mu_3$. *Let* a_1, a_2 *and* a_3 *denote the corresponding correct actions. We have*

- *If* $\mu_3 > \frac{1}{e^{1/\lambda}+2}$, *then* $p^\star_{a_i}(\mu_b, \mu_g) = \frac{\mu_i\left(e^{1/\lambda}+2\right)-1}{e^{1/\lambda}-1}$ *for* $i \in \{1,2,3\}$.
- *If* $\mu_3 < \frac{1}{e^{1/\lambda}+2} \& \mu_1 < \mu_2 e^{1/\lambda}$, *then* $p^\star_{a_i}(\mu_b, \mu_g) = \frac{\frac{\mu_i}{\mu_1+\mu_2}\left(e^{1/\lambda}+1\right)-1}{e^{1/\lambda}-1}$ *for* $i \in \{1,2\}$ & $p^\star_{a_3}(\mu_b, \mu_g) = 0$.
- *If* $\mu_3 < \frac{1}{e^{1/\lambda}+2} \& \mu_1 > \mu_2 e^{1/\lambda}$, *then* $p^\star_{a_1}(\mu_b, \mu_g) = 1$, $p^\star_{a_2}(\mu_b, \mu_g) = p^\star_{a_3}(\mu_b, \mu_g) = 0$

The first point of Proposition 5 states that if the lowest prior belief μ_3 is higher than a certain threshold, then the DM chooses among all three actions at optimality. Else, the DM chooses between two actions if μ_1 and μ_2 take sufficiently close values (second point of the proposition). Otherwise, the DM only chooses one state (third point). In this sense, Proposition 5 characterizes the possible consideration sets of alternatives from which the DM chooses, as a function of priors μ_g and μ_b.

Figure 14 illustrates these sets for two different information cost values. For instance, when both μ_g and μ_b are sufficiently low, the DM chooses $a = o$ without eliciting any signal. The regions labeled as "Only a" correspond to situations where the DM directly chooses action a and disregards the two others a priori. In these cases, the DM does not process any information. Similarly, the DM rules out the moderate state a priori and chooses $a = y$ and $a = n$ with positive probability (i.e., forms a consideration set consisting only the good and bad states) if her belief toward the moderate state is sufficiently weak (see the region labeled as "Both y&n"). Only in the region labeled as "All" does the DM consider all possible states at the same time. As indicated in Figure 14, this region shrinks when λ increases as the DM becomes less willing to process information for weaker prior beliefs.

Using Proposition 5 and following an approach similar to our base model, we can characterize the DM's decision accuracy, error probabilities and cognitive effort. These are detailed at the end of this section (see Appendix §E.1). In the following, we focus on how the presence of the machine impacts these metrics.

Upon a positive machine input, the DM rules out the bad state and updates her prior as $\mu_b^+ = 0$ and $\mu_g^+ = \frac{\mu_g}{P(X_1=+)}$. Similarly, upon a negative machine input, the DM rules out the good state with posterior beliefs $\mu_b^- = \frac{\mu_b}{P(X_1=-)}$ and $\mu_g^- = 0$. That is, the DM has always two states to consider in the presence of the machine: the moderate state and either the good or the bad state depending on the machine's input. Nonetheless, the machine always improves the DM's decision accuracy and expected value as we show in the next proposition.

PROPOSITION 6. *For any* $\lambda > 0$, *we have* $A^\star_m \geq A^\star$ *and* $V^\star_m \geq V^\star$.

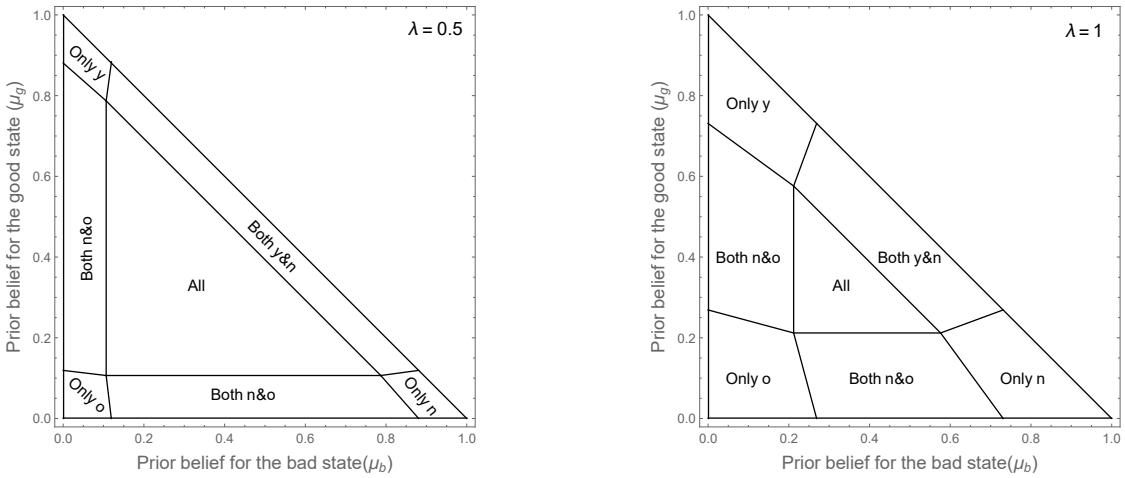


Figure 14 Consideration set formation in $\mu_b \times \mu_g$ space

Although the machine always increases overall accuracy, the machine may increase certain error types and induce the DM to exert more cognitive effort as in our base model. To explore this, we can follow our previous approach, and compare these different metrics with and without machine prediction using our characterizations in Appendix E.1. In particular, this consists in studying different scenarios, which depend on the different prior and posterior beliefs. However, the number of possible scenarios to consider grows exponentially with the number of possible states (from six possible cases in the base model with two states to 63 with three states.)[9] Although technically doable, exploring all these situations is tedious and of limited research interest. Instead, we focus on the situation in which the machine reduces uncertainty in a symmetric manner, i.e. where the DM's prior and posterior beliefs are symmetric. Specifically, we take $\mu_g = \mu_b = \mu < 0.5$ and $\pi(+,-) = \pi(-,+) = 1/2 - \mu$, so that the DM's posterior beliefs become $\mu_g^+ = \mu_b^- = 2\mu$ and the likelihood of a machine outcome is given by $P(X_1 = +) = P(X_1 = -) = 0.5$. We next characterize the impact of the machine on the DM's decision errors and cognitive effort in this set-up.

PROPOSITION 7. *For any given $\lambda > 0$, $\alpha_m^\star > \alpha^\star$ if $\frac{1}{2}\frac{1}{e^{1/\lambda}+1} < \mu < \mu_{fp}^*$ and $\alpha_m^\star \leq \alpha^\star$ otherwise. Furthermore, $\mu_{fp}^* < 1/3$, $\alpha_m^\star = \beta_m^\star$ and $\alpha^\star = \beta^\star$*

Note that the false positive and negative errors are always equal, whether the machine is present or not ($\alpha_m^\star = \beta_m^\star$ and $\alpha^\star = \beta^\star$). This is because the DM's prior and posterior beliefs are symmetric in this set-up. In essence, Proposition 7 states that the machine increases the DM's false positive (and negative) errors if she sufficiently favors the moderate state ($\mu_m > 1 - 2\mu_{fp}^* > 1/3$) a priori. This increase in the decision errors is actually offset by a decrease in *false moderate* errors, so that the machine always improves overall accuracy per Proposition 6.

[9] The inclusion of a new state gives rise to 7 different cases as depicted in Figure 14 when the human DM is alone (as opposed to 3 in our base model which are "Only y", "Only n" and "Both y&n"). In the presence of the machine, we have 9 cases -3 cases for each of the two possible posterior beliefs given the machine's prediction on $X1$ (this is since the machine always eliminates one of the states for the DM). This requires studying up to 7x9=63 cases for a full-fledged analysis of the problem. By contrast, there are 3x3=9 cases in our base model - as there are only 3 cases to consider with the machine since $\mu^- = 0$. Some of these cases can further be ruled out since $\mu^+ > \mu$, which yields 6 possible cases for our base model as illustrated in Theorems 2, 3 and 4.

PROPOSITION 8. *For any given $\lambda > 0$, $C_m^\star > C^\star$ if $\frac{1}{2}\frac{1}{e^{1/\lambda}+1} < \mu < \mu_e^*$ and $C_m^\star \leq C^\star$ otherwise. Furthermore, $\mu_e^* < 1/3$.*

Proposition 8 shows that the machine increases the cognitive effort the DM exerts when she sufficiently favors the moderate state. In this case, the task difficulty increases with any machine input. To see this, recall that the DM without the machine tries to distinguish three states with initial belief $\mu_b = \mu_g = \mu$ and $\mu_m = 1 - 2\mu$. When the machine gives a positive (resp. negative) signal, the DM deals with two states with posterior $\mu_g = 2\mu$ (resp. $\mu_b = 2\mu$) and $\mu_m = 1 - 2\mu$. When μ is sufficiently small, distinguishing the remaining two states becomes more difficult for the DM.

Overall, Proposition 7 and Proposition 8 establish that our main results for the base model are robust to decision settings where the machine reduces uncertainty in a symmetric manner.

E.1. Accuracy and Cognitive Effort in the Presence of Three States

Accuracy and Decision Errors We denote by $p^\star_{a|\omega}(\mu_b, \mu_g)$ the posterior probability that the DM chooses action a given state ω. Accuracy is defined as the sum of joint probability of choosing $a = y$ in the good state, choosing $a = n$ in the bad state and choosing $a = o$ in the moderate state. These joint probabilities can be written in terms of the DM's optimal choice probabilities and corresponding posteriors. From Theorem 1 in Caplin et al. (2019), the DM's optimal posteriors are

$$p^\star_{\omega_j|a_i} = \begin{cases} \frac{e^{1/\lambda}}{e^{1/\lambda}+2} & \text{for } i = j \\ \frac{1}{e^{1/\lambda}+2} & \text{otherwise} \end{cases} \text{ if } \mu_3 > \frac{1}{e^{1/\lambda}+2}$$

$$p^\star_{\omega_j|a_i} = \begin{cases} \frac{e^{1/\lambda}(\mu_1+\mu_2)}{e^{1/\lambda}+1} & \text{for } i = j \text{ and } j \leq 2 \\ \frac{(\mu_1+\mu_2)}{e^{1/\lambda}+1} & \text{for } i \neq j \text{ and } j \leq 2 \\ \mu_j & j = 3 \end{cases} \text{ if } \mu_3 < \frac{1}{e^{1/\lambda}+2} \text{ and } \mu_1 < \mu_2 e^{1/\lambda}$$

$$p^\star_{\omega_j|a_i} = \mu_j \text{ if } \mu_3 < \frac{1}{e^{1/\lambda}+2} \text{ and } \mu_1 > \mu_2 e^{1/\lambda}$$

Plugging in these and after some algebra, we obtain the accuracy function as

$$\begin{aligned} A^\star(\mu_b, \mu_g) &= p^\star_{g|y} p^\star_y(\mu_b, \mu_g) + p^\star_{b|n} p^\star_n(\mu_b, \mu_g) + p^\star_{m|o} p^\star_o(\mu_b, \mu_g) \\ &= \begin{cases} \mu_1 & \text{if } \mu_3 < \frac{1}{e^{1/\lambda}+2} \text{ and } \mu_1 > \mu_2 e^{1/\lambda} \\ \frac{e^{1/\lambda}(\mu_1+\mu_2)}{e^{1/\lambda}+1} & \text{if } \mu_3 < \frac{1}{e^{1/\lambda}+2} \text{ and } \mu_1 < \mu_2 e^{1/\lambda} \\ \frac{e^{1/\lambda}}{e^{1/\lambda}+2} & \text{if } \mu_3 > \frac{1}{e^{1/\lambda}+2} \end{cases} \end{aligned} \tag{17}$$

The resulting characterization has many similarities with our base model with two actions and two states. Note that when only one option is selected, then DM's accuracy is just her prior for that state which, in fact, will be the one with the strongest prior belief. When all options are selected, expected accuracy does not depend on prior belief. Similar to our base model, this is related to the symmetry of DM's payoffs so that her posterior beliefs are also symmetric. On the other hand, when the DM selects only two options in optimality, expected accuracy is just the scaled-down version of the accuracy in our base model with the likelihood of both selected states (i.e., $(\mu_1 + \mu_2)$). Note that $e^{1/\lambda}/\left(e^{1/\lambda}+1\right)$ is the expected accuracy in our base model when both options are selected. Due to this scaling, expected accuracy in this case depends on the DM's prior belief.

In a similar manner, the error probabilities (respectively false positive, false negative and false moderate) can be computed by plugging in choice probabilities and optimal posteriors as follows:

$$\alpha^\star(\mu_b,\mu_g) = p_y^\star(\mu_b,\mu_g)\left(1-p_{g|y}^\star\right)$$
$$\beta^\star(\mu_b,\mu_g) = p_n^\star(\mu_b,\mu_g)\left(1-p_{g|n}^\star\right)$$
$$\gamma^\star(\mu_b,\mu_g) = p_o^\star(\mu_b,\mu_g)\left(1-p_{g|o}^\star\right)$$

Cognitive Effort When there are three states, the entropy as a function of μ_g and μ_b is

$$H(\mu_b,\mu_g) = -\mu_g\log\mu_g - \mu_b\log\mu_b - (1-\mu_g-\mu_b)\log(1-\mu_g-\mu_b).$$

The DM's cognitive effort is then defined as

$$C^\star(\mu_b,\mu_g) = \lambda\left[H(\mu_b,\mu_g) - p_y^\star(\mu_b,\mu_g)\,H\left(p_{g|y}^\star,p_{b|y}^\star\right) - p_y^\star(\mu_b,\mu_g)\,H\left(p_{g|n}^\star,p_{b|n}^\star\right) - p_y^\star(\mu_b,\mu_g)\,H\left(p_{g|o}^\star,p_{b|o}^\star\right)\right]$$

which reduces (after some algebra) to

$$C^\star(\mu_b,\mu_g) = \begin{cases} 0 & \text{if } \mu_3 < \frac{1}{e^{1/\lambda}+2} \text{ and } \mu_1 > \mu_2 e^{1/\lambda} \\ \lambda\left[H(\mu_b,\mu_g) - (\mu_1+\mu_2)\,\varphi(\lambda) - H(\mu_1+\mu_2,0)\right] & \text{if } \mu_3 < \frac{1}{e^{1/\lambda}+2} \text{ and } \mu_1 < \mu_2 e^{1/\lambda} \\ \lambda\left[H(\mu_g,\mu_b) - \varphi^{(2)}(\lambda)\right] & \text{if } \mu_3 \geq \frac{1}{e^{1/\lambda}+2} \end{cases} \tag{18}$$

where

$$\varphi(\lambda) = \ln\left(e^{1/\lambda}+1\right) - \frac{1}{\lambda}\frac{e^{1/\lambda}}{e^{1/\lambda}+1}, \tag{19}$$

and

$$\varphi^{(2)}(\lambda) = -\frac{1}{\lambda}\frac{e^{\frac{1}{\lambda}}}{e^{\frac{1}{\lambda}}+2} + \ln\left(e^{\frac{1}{\lambda}}+2\right). \tag{20}$$

When the DM's prior is sufficiently strong toward a particular state, she does not process information and chooses the corresponding action. Then her cognitive effort is 0. When the DM chooses to process information about all states then her expected cognitive effort is in a very similar structure as our base model with two states. In particular, it is directly proportional to the difference between DM's prior entropy and expected posterior entropy given by $\varphi^{(2)}(\lambda)$. Note that it is independent of DM's prior belief and increasing in information cost λ. Lastly, when DM chooses to select two options and disregards one, then her expected posterior entropy depends on her prior belief.

Appendix F: The Impact of Restricting Information Sources for the DM

An information acquisition model for human decision maker which predicts that a free and perfectly accurate information (X_1 in our model) deteriorates her overall accuracy is not credible, at least in our view. This is, however, what a classical sequential hypothesis testing model as in DeGroot (1962) would predict in our context. In this section we illustrate this point. The key difference that the sequential testing introduces is that it limits the choice of signals (i.e. tests) that the DM can elicit.

Assume that at each period the DM may decide to conduct an imperfect test which provides a positive or negative signal about the true state. Each test costs K to the DM. There are infinite number of tests, but we assume that the DM decides optimally when to stop the search and commit to a decision. This corresponds to a standard optimal-stopping problem. To formulate this, we denote the machine signal at

period k as $Z_k \in (+,-)$. We assume that the accuracy of the test is exogenous and known to the DM a priori. Assume that the test gives a signal with a true positive rate $\phi = P(Z_k = +|\omega = g)$ and false positive rate $\psi = P(Z_k = +|\omega = b)$. We denote the DM's prior belief that the state is good at period k as $\mu_k = P(\omega = g)$. If the DM decides to conduct the test, she updates her prior belief as following the Bayes' rule:

$$\mu_{k+1} = \begin{cases} \frac{\alpha\mu_k}{\alpha\mu_k+\beta(1-\mu_k)} & \text{if } Z_k = + \\ \frac{(1-\alpha)\mu_k}{(1-\alpha)\mu_k+(1-\beta)(1-\mu_k)} & \text{if } Z_k = - \end{cases}.$$

As in our base model, the DM aims to maximize her accuracy net of her total search cost. This is an optimal-stopping problem. We can write the corresponding DP formulation as

$$V_k(\mu_k) = \max\{E_{Z_k}[V_{k+1}(\mu_{k+1})] - K, \max\{\mu_k, 1-\mu_k\}\}$$
$$V_T(\mu_T) = \max\{\mu_T, 1-\mu_T\}$$

where T is the terminal period. Here $E_{Z_k}[V_{k+1}(\mu_{k+1})] - \lambda$ is the DM's expected utility if she chooses to continue sampling where $P(Z_k = 1) = \alpha\mu_k + \beta(1-\mu_k)$. If she decides to choose an action and stop observing the signal, she obtains $\max\{\mu_k, 1-\mu_k\}$ by deciding based on her prior μ_k.

Let us denote DM's optimal information acquisition decision at period k as u_k^*, which can be written as

$$u_k^* = \begin{cases} 1 & \text{if } \max\{\mu_k, 1-\mu_k\} < E_Z[V_{k+1}(\mu_{k+1})] - K. \\ 0 & \text{otherwise} \end{cases}$$

where 1 is continue sampling and 0 is stop sampling. Accordingly, we can write the DM's decision accuracy function as

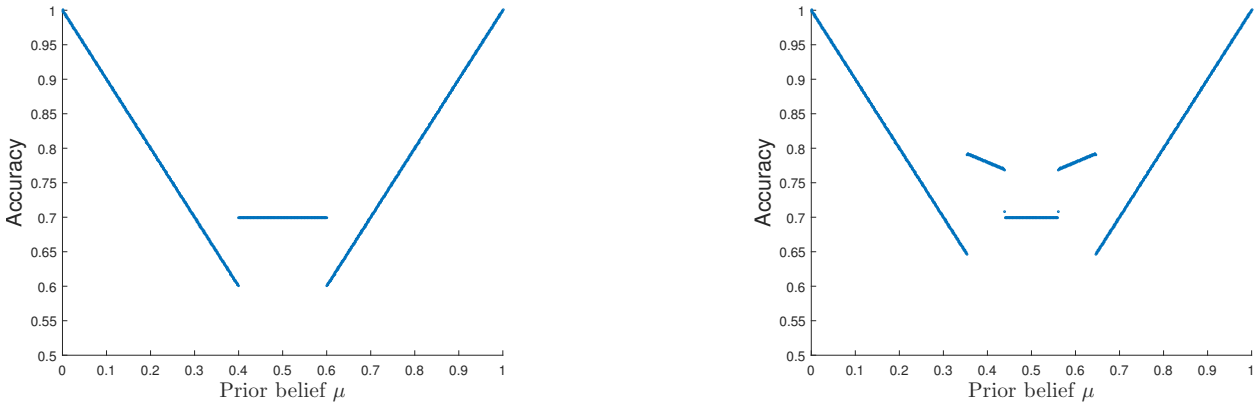
$$A_k(\mu_k) = \begin{cases} P(Z_k = 1) A_{k+1}\left(\frac{\alpha\mu_k}{\alpha\mu_k+\beta(1-\mu_k)}\right) + P(Z_k = 0) A_{k+1}\left(\frac{(1-\alpha)\mu_k}{(1-\alpha)\mu_k+(1-\beta)(1-\mu_k)}\right) & \text{if } u_k^* = 1 \\ \max\{\mu_k, 1-\mu_k\} & \text{if } u_k^* = 0 \end{cases}$$
$$A_T(\mu_k) = \max\{\mu_T, 1-\mu_T\}.$$

As in our base model, we assume that the machine can reveal X_1 at no cost. Depending on the DM's updated prior upon the machine information, she can conduct further tests as explained above. We compare the DM's overall accuracy in these two cases (with and without the machine).

As a numerical example take $\phi = 0.3$, $\psi = 0.7$ and $K = 0.1$. Numerically solving the corresponding optimal stopping problem yields the accuracy function depicted in Figure 15a for the DM. In this case, she chooses not to conduct any test if her prior for the good state μ is less than 0.4 or higher than 0.6. When her prior belief is sufficiently close to 0.5 ($\mu \in [0.4, 0.6]$), she chooses to conduct a single test and decides accordingly which yields an accuracy level of 0.7. Note that the DM's accuracy is neither convex nor continuous in prior belief unlike our model that is based on the rational inattention framework (please see Figure 2a).

To see how the machine decreases the DM's accuracy in this scenario, take, for instance, $\mu = 0.5$. This yields an accuracy level of $A = 0.7$ per Figure 15a. Assume further that the DM's posterior belief is $\mu^+ = 0.6$ when the machine provides a positive signal. This yields an ex-post accuracy of 0.6. Since the probability of a positive signal by the machine is $P(X_1 = +) = 0.5/0.6 = 5/6$, the DM's expected accuracy with the machine is $A_m = 5/6 * 0.6 + 1/6 * 1 = 0.667$ which is less than 0.7, the DM's accuracy without the machine.

Figure 15b plots DM's accuracy function for a slightly lower test cost $K = 0.08$. In this case, the DM chooses to conduct more than one test when she is sufficiently uncertain as information cost is lower. Similarly, we



(a) $K = 0.1$ (b) $K = 0.08$

Figure 15 DM's decision accuracy as a function of prior belief ($\alpha = 0.7$, $\beta = 0.3$)

can also find parameters in this scenario where machine can strictly decrease the DM's overall expected accuracy.

These simple counter-examples show that if the DM is able to elicit costly information from a *restrictive* set of signal sources, it is possible that a free and accurate machine information may actually reduce her overall accuracy.

Recent ESMT Working Papers

	ESMT No.
Mapping Markush Stefan Wagner, ESMT European School of Management and Technology Christian Sternitzke, Sternitzke Ventures UG Sascha Walter, University of Würzburg	22-05
Decertification in quality-management standards by incrementally and radically innovative organizations Joseph A. Clougherty, University of Illinois at Urbana-Champaign Michał Grajek, ESMT European School of Management and Technology	22-04
Do decision makers have subjective probabilities? An experimental test David Ronayne, ESMT European School of Management and Technology Roberto Veneziani, Queen Mary University of London William R. Zame, University of California at Los Angeles	22-03
Is your machine better than you? You may never know. Francis de Véricourt, ESMT Berlin Huseyin Gurkan, ESMT Berlin	22-02
Mismanaging diagnostic accuracy under congestion Mirko Kremer, Frankfurt School of Finance and Management Francis de Véricourt, ESMT Berlin	22-01