Making Information More Valuable

Mark Whitmeyer*

December 15, 2023

Abstract

We study what changes to an agent's decision problem increase her value for information. We prove that information becomes more valuable if and only if the agent's reduced-form payoff in her belief becomes more convex. When the transformation corresponds to the addition of an action, the requisite increase in convexity occurs if and only if a simple geometric condition holds, which extends in a natural way to the addition of multiple actions. We apply these findings to two scenarios: a monopolistic screening problem in which the good is information and delegation with information acquisition.

Keywords: Expected Utility, Selling Information, Information Acquisition, Delegation
JEL Classifications: D81; D82; D83

---

*Arizona State University. Email: mark.whitmeyer@gmail.com. I am grateful to Simon Board, Costas Cavounidis, Gregorio Curello, Dana Foarta, Rosemary Hopcroft, Vasudha Jain, Doron Ravid, Eddie Schlee, Ludvig Sinander, Bruno Strulovici, Can Urgun, Tong Wang, Joseph Whitmeyer, Tom Wiseman, Renkun Yang, Kun Zhang, and various seminar and conference audiences for their feedback. I also thank the editor, Emir Kamenica, and four anonymous referees for their useful suggestions. This paper was formerly titled "Flexibility and Information."

When action grows unprofitable, gather information;
when information grows unprofitable, sleep.

Ursula Le Guin, *The Left Hand of Darkness*

1 Introduction

It is more-or-less accepted that in typical environments, economic agents are risk averse. All else equal, a rational expected-utility maximizer prefers less uncertainty to more. Similarly, rational Bayesians are information loving: more information always weakly improves a decision-maker's payoff. Of course, some agents dislike risk more than others, and Pratt (1964) formulates a compelling way to characterize precisely what it means for one agent to be more risk-averse than another.

Some agents value information more than others, so, in a similar vein, it is natural to seek a way to rank agents' love for information. When can we say that one agent values information more than another? Equivalently, suppose we alter an agent's decision problem. What sorts of modifications increase her value for information?

Pratt's way of comparing agents' risk aversion is relatively detail-free. Agent 1 is more risk averse than agent 2 if, *for any lottery*, agent 1's certainty equivalent–the minimum sure-thing payoff she would accept in lieu of the lottery–is smaller than agent 2's. In specifying what it means for agent 1 to value information more than agent 2, we take a similarly broad approach. Here, we require that *any* information be more valuable to agent 1 than agent 2, *no matter the prior*.

As revealed by Pratt (1964), relative **concavity** is the key to comparing agents' aversion to risk. Agent 1 is more risk averse than agent 2 if her utility function u_1 is a concave transformation of agent 2's utility u_2: $u_1 = \phi \circ u_2$ for some strictly monotone concave function ϕ. As we discover in this paper, relative **convexity**, of a different variety, distinguishes agents' comparative love of information. Agent 1 values information more than agent 2 if the difference in the agents' value functions[1] $V_1 - V_2$ is convex.

[1] An agent's value function, $V(x)$, is her maximal expected payoff at any belief $x \in \Delta(\Theta)$ as a function of that belief, obtained by plugging in an optimizing action.

After identifying the connection between $(V_1 - V_2)$'s convexity and an agent's comparative love of information, we turn our attention to the value functions themselves. What modifications to an agent's decision problem result in an increase in convexity? One natural way to alter an agent's decision problem is by giving her an additional action. How does increased flexibility–a greater capacity to adapt her behavior to new information–change an agent's value for information?

In §4.2, we carry this analysis further by giving the agent not just one but potentially multiple actions. Next, in §4.3, we remove actions. We then leave the set of actions unchanged (§4.4), but instead scale the agent's utility. This allows us to speak to the effects of repetition and aggregate risk on the value of information.

In the leading application of our main result–the addition of a single action (§4.1)–we uncover a simple geometric condition necessary and sufficient for the requisite increased convexity of the agent's value function. An iterative version of this condition also guarantees an increase in convexity when multiple actions are added. Moreover, although the condition is not necessary for an increase in convexity when multiple actions are added, any failure of necessity is not robust–perturbing the utilities from the new actions slightly will destroy the increase in convexity.

Perhaps unexpectedly, we discover that unless all of the remaining actions or all of the removed actions were initially dominated, taking away actions can never lead to a higher value for information. That is, it is only an elimination that results in a totally new decision problem (in effect) or the exact same decision problem, that can lead to an increase in an agent's value for information.

1.1 Motivating Example

The question under study has significant practical relevance. The job of a regulator is to enact policies that modify the incentives of agents in some environment. This typically entails the addition or subtraction of actions: there are contracts that an insurer may not offer, assets that an investment firm may not sell, and limits to how many fish a trawler may catch.[2] Insurers themselves change agents' payoffs by reducing their risk, flattening their payoffs. Firms do the opposite with their workers: bonus schemes tied to a worker's performance make her payoff steeper and more sensitive to randomness.

Consider for instance an insurance provider dictating what treatments it will cover; *viz.*, what procedures a doctor may conduct. For simplicity, suppose there are three conditions a patient with an injured hand may have–three states of the world. In one state, state 0, the injury is just a sprain; in another, state 1, a bone is broken but not displaced; and in state 2, the fracture is displaced.

Suppose first the doctor may only offer one treatment: place a cast on the hand (action c). Accordingly, she has two possible actions, do nothing (action n), which is uniquely optimal if the injury is just a sprain; or cast the hand, which is uniquely optimal if the hand is broken. This decision problem is represented in Figure 1a: point (x, y) specifies the respective probabilities (beliefs) that the bone is broken but not displaced or broken and displaced. Accordingly, the blue region is the region of probabilities in which n is optimal; and the red region are those probabilities for which c is optimal.

Let us now consider two possible new treatments afforded to the doctor. Suppose the provider now covers surgery (action s). This is relatively high-risk and is only optimal if the doctor is confident the bone is broken and displaced. Figure 1b represents this new decision problem: s is optimal if and only if the doctor's belief is in the purple region. On the other hand, suppose the provider instead allows a conservative treatment consisting of stretching and rehabilitating exercises (action r). This is better than nothing in the case of a fracture, but is inferior to rest for sprains. This scenario is Figure 1c, where r is optimal for beliefs in the black region.

Which of these new options, if either, does not dampen the doctor's enthusiasm for information, regardless of her prior or what that information may be? As we discover in this paper, the answer is simple, only the former of the two potential new procedures, surgery, makes information more valuable. Indeed, suppose that a sprain and a break are equally likely. With only the initial two treatments to choose from, the doctor strictly

[2] The verdicts that can be handed down in criminal cases are also legislated, so our results also speak to what kinds of verdicts improve incentives for information acquisition (cf. Siegel and Strulovici (2020)).

benefits from any information. If we gave the doctor the conservative option, this would clearly no longer be true: any information that doesn't move her beliefs much is now worthless, as the conservative treatment remains optimal at those beliefs. In contrast, the surgery option makes information weakly more valuable.

The crucial difference between the two potential new actions is that the new action is **refining:** only the region of beliefs in which doing nothing is optimal shrinks. In contrast, the conservative treatment partially replaces each pre-existing treatment.

2 The Model

There is a grand set of actions $\mathcal{A}$. Our protagonist is a decision maker, an agent who initially possesses a compact set of actions $A \subseteq \mathcal{A}$. There is an unknown state of the world θ, which is drawn according to some full-support prior μ from some finite set of states Θ. Initially, the agent has some continuous utility function $u\colon \mathcal{A} \times \Theta \to \mathbb{R}$.

$\mathcal{D} := (u, A, \Theta)$ denotes the agent's Initial Decision Problem. We are studying the effect of a transformation of the decision problem on the agent's value for information.[3] To that end, $\hat{\mathcal{D}} := \left(\hat{u}, \hat{A}, \Theta\right)$ denotes the agent's Transformed Decision Problem. Here are a few leading examples of such transformations:

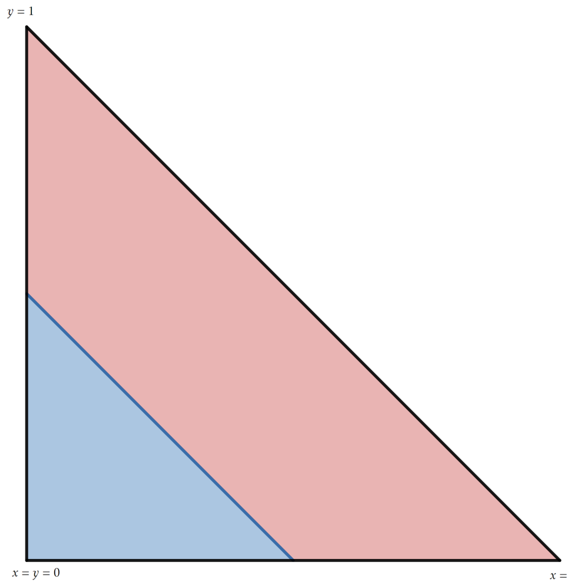
Scenario 1. Becoming More Flexible. A is finite. In $\hat{\mathcal{D}}$, the agent's utility function remains unchanged, $\hat{u} = u$; and her new set of actions is $\hat{A} := A \cup \{\hat{a}\}$ for some $\hat{a} \in \mathcal{A} \setminus A$.

Scenario 2. Becoming Much More Flexible. Again, A is finite; and in $\hat{\mathcal{D}}$, the agent's utility function stays the same, $\hat{u} = u$. Now, her new set of actions is $\hat{A} := A \cup B$ for some additional finite set of actions $B \in \mathcal{A} \setminus A$.

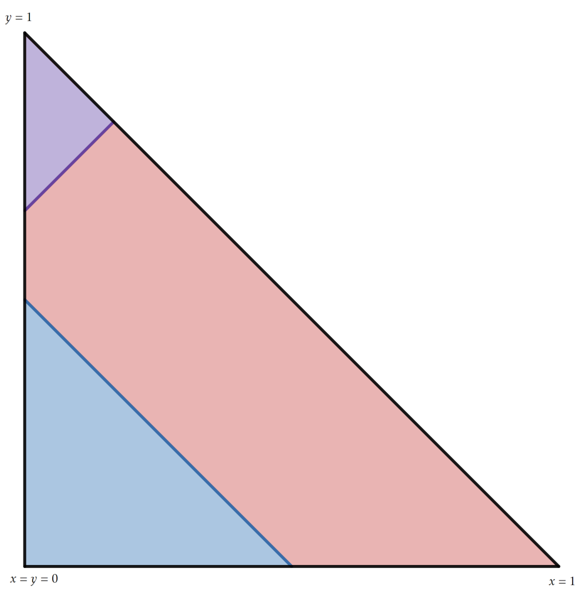
Scenario 3. Becoming Much Less Flexible. Yet again, A is finite and $\hat{u} = u$. Now, however, the agent loses actions, moving from $\mathcal{D}$ to $\hat{\mathcal{D}}$: $\emptyset \neq \hat{A} \subset A$.

Scenario 4. Transforming the Agent's Utility Function. In $\hat{\mathcal{D}}$, $\hat{u} = \phi \circ u$ for some strictly monotone, continuous ϕ; and her new set of actions is unaltered: $\hat{A} = A$.

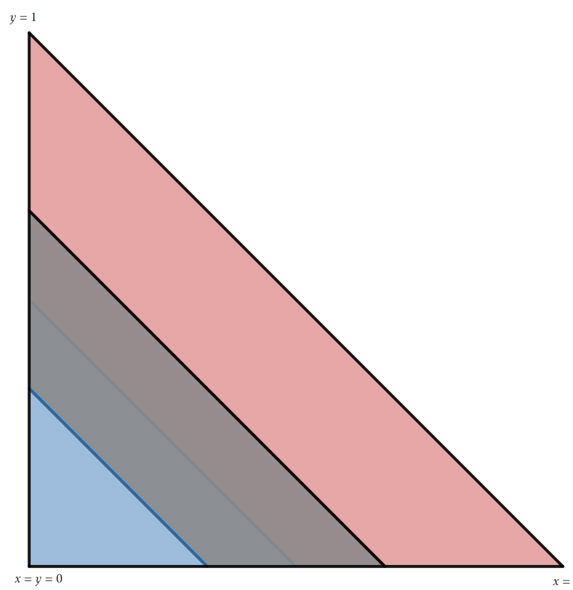
[3] Equivalently, we are comparing two different agents' values for information.



(a) Choosing between c and n



(b) Choosing between c, n, and s



(c) Choosing between c, n, and r

Figure 1: The subdivision of the 2-simplex corresponding to the motivating example. There are three states $\{0, 1, 2\}$, with $x := \mathbb{P}(1)$ and $y := \mathbb{P}(2)$.

$\Delta(\Theta)$ denotes the simplex of probability distributions over Θ and $\operatorname{int}\Delta(\Theta)$ its interior.[4] When the agent acquires information, she does so by observing the realization of an experiment, stochastic map $\pi\colon \Theta \to \Delta(S)$, where S is a compact set of signal realizations. Equivalently (Kamenica and Gentzkow (2011)), she obtains a Bayes-plausible distribution over posterior beliefs (posteriors) $\Phi \in \mathcal{F}(\mu) \subseteq \Delta(\Delta(\Theta))$.[5]

Given an initial decision problem $\mathcal{D}$, the agent's Value Function, in belief $x \in \Delta(\Theta)$, is

$$V(x) := \max_{a\in A} \mathbb{E}_x u(a,\theta).$$

$\hat{V}$ is the analogous object in the transformed decision problem, and both functions are convex. There are two natural ways to understand an increase in an agent's value for information.

1. Exogenous information. We say that A Transformation Generates a Greater Value for Information if

$$\mathbb{E}_\Phi \hat{V}(x) - \hat{V}(\mu) \geq \mathbb{E}_\Phi V(x) - V(\mu),$$

for all $\Phi \in \mathcal{F}(\mu)$ and $\mu \in \operatorname{int}\Delta(\Theta)$. A transformation generates a greater value for information if the resulting expected payoff as a result of obtaining information is greater in the transformed decision problem than in the initial decision problem, no matter the information (for any Bayes-plausible Φ) and no matter the prior (any $\mu \in \operatorname{int}\Delta(\Theta)$).

2. Endogenous information. Our second way of interpreting the value of information gives the agent greater control over information acquisition. Now, the experiment is not exogenous but instead an endogenous choice of the agent. Given an initial decision problem $\mathcal{D}$ and a prior μ, in the agent's flexible information acquisition problem, she solves

$$\max_{\Phi\in\mathcal{F}(\mu)} \int_{\Delta(\Theta)} V(x)\, d\Phi(x) - D(\Phi), \tag{$\star$}$$

where D is a uniformly posterior-separable cost functional,[6] that for which there exists a

[4] For a set Y, $\operatorname{int} Y$ denotes its relative interior.

[5] A distribution Φ is Bayes-plausible if it is supported on a subset of $\Delta(\Theta)$ and $\mathbb{E}_\Phi(x) = \mu$.

[6] This family of costs is introduced in Caplin, Dean, and Leahy (2022).

strictly convex function $c\colon \Delta(\Theta) \to \mathbb{R}$ such that

$$D(\Phi) = \int_{\Delta(\Theta)} c(x)\, d\Phi(x) - c(\mu).$$

Similarly, in the transformed decision problem $\hat{\mathcal{D}}$, the agent solves

$$\max_{\Phi \in \mathcal{F}(\mu)} \int_{\Delta(\Theta)} \hat{V}(x)\, d\Phi(x) - D(\Phi), \tag{$\hat{\star}$}$$

We say that A Transformation Does Not Generate Less Information Acquisition if for any prior $\mu \in \operatorname{int}\Delta(\Theta)$, UPS cost functional D, and solution to the agent's information acquisition problem in the initial decision problem (Problem $\star$), Φ^*, there exists an solution to the agent's information acquisition problem in the transformed decision problem (Problem $\hat{\star}$), $\hat{\Phi}^*$, that is not a strict mean-preserving contraction (MPC) of Φ^*.[7] A transformation does not generate less information acquisition if for any optimal information acquisition strategy in $\mathcal{D}$, there is an optimal information acquisition strategy in $\hat{\mathcal{D}}$ in which the agent does not acquire strictly less information.

2.1 Geometric Preliminaries

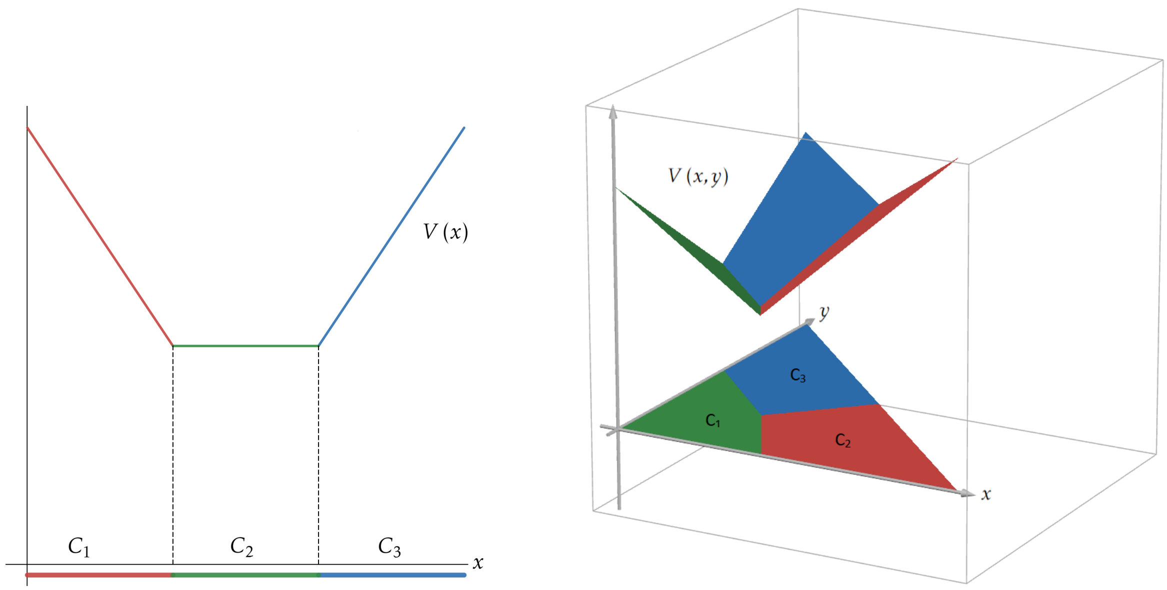
When A is finite, we say an action $a_i \in A$ is not Weakly Dominated, or is Undominated, if there exists a belief $x \in \Delta(\Theta)$ such that

$$\mathbb{E}_x u(a_i, \theta) > \max_{a \in A \setminus \{a_i\}} \mathbb{E}_x u(a, \theta).$$

If A is finite, the agent's value function, V, is piecewise affine, and its graph is a polyhedral surface in $\mathbb{R}^n$, where n is the number of states. Associated with V is its projection onto $\Delta(\Theta)$, which yields a finite collection C of polytopes of full dimension C_i ($i = 1, \dots, m$), where m is the number of undominated actions in A. Formally, for each undominated $a_i \in A$ ($i = \{1, \dots, m\}$),

$$C_i := \{x \in \Delta(\Theta) \mid \mathbb{E}_x u(a_i, \theta) = V(x)\} = \left\{x \in \Delta(\Theta) \;\middle|\; \mathbb{E}_x u(a_i, \theta) \geq \max_{a \in A \setminus \{a_i\}} \mathbb{E}_x u(a, \theta)\right\}.$$

[7]For distributions P and Q supported on a compact, convex subset, X, of a vector space, P is an MPC of Q if $\int \phi dP \leq \int \phi dQ$ for all convex functions $\phi\colon X \to \mathbb{R}$. P is a strict MPC of Q if P is an MPC of Q but Q is not an MPC of P. Q is a mean-preserving spread (MPS) of P if P is an MPC of Q.



(a) V and C, two states.

(b) V and C, three states.

Figure 2

By construction, action a_i is optimal for any belief $x \in C_i$ and uniquely optimal for any belief $x \in \operatorname{int} C_i$. The collection C is a Regular Polyhedral Subdivision (henceforth, Subdivision) of $\Delta(\Theta)$. Figure 2 illustrates two pairs of value functions and subdivisions when the agent has three actions. Each C_i is a Cell of C.

If the set of actions in the transformed decision problem, $\hat{A}$, is also finite, the new value function $\hat{V}$, itself, has a corresponding subdivision, $\hat{C}$. There is a natural way of comparing subdivisions that is useful for our purposes: a subdivision $P = \{P_1, \dots, P_l\}$ is Finer than (or Refines) a subdivision $Q = \{Q_1, \dots, Q_m\}$ if for each $j \in \{1, \dots, l\}$, there exists $i \in \{1, \dots, m\}$ such that $P_j \subseteq Q_i$ (Lee and Santos (2017)). We write this $P \succeq Q$ ($\succ$ when the relation is strict). In anticipation of our later results, we note a tight connection between the refinement order and relative convexity of the value functions:

Lemma 2.1. *$\hat{V} - V$ is convex only if $\hat{C} \succeq C$.*

Proof. Please visit Appendix A.1. ∎

It is easy to see that even if the agent only gains one additional action, when we compare the subdivisions C and $\hat{C}$, we can say nothing in general about their relationship in the finer-than partial order. For instance, if the new action, $\hat{a}$, strictly dominates all of the actions in A, $\hat{C}$ has a single cell, $\Delta(\Theta)$, so $C \succeq \hat{C}$ (Figure 3c). Moreover, the new action can be such that C and $\hat{C}$ are incomparable (Figure 3a) or such that $\hat{C} \succeq C$ (Figure 3b).

The last outcome is special. We say that a new action $\hat{a}$ is Refining if $\hat{a}$ is not weakly dominated in $A \cup \{\hat{a}\}$ and $\hat{C} \succeq C$. That is, there exists $x \in \Delta(\Theta)$ for which $\mathbb{E}_x u(\hat{a}, \theta) > V(x)$ and

$$\{x \in \Delta(\Theta) \mid \mathbb{E}_x u(\hat{a}, \theta) \geq V(x)\} \subseteq C_i,$$

for some $C_i \in C$. Refining actions are those that are good–uniquely optimal in at least one state of the world–but not too good–it is only one undominated action in A whose region of unique optimality shrinks as a result of adding the refining action.

3 Making Information More Valuable

In this section, we state and prove the main result of the paper:

Theorem 3.1. *A three-way equivalence holds:*

$\hat{V} - V$ *is convex.*	$\Leftrightarrow$	*A transformation does not generate less information acquisition.*	$\Leftrightarrow$	*A transformation generates a greater value for information.*

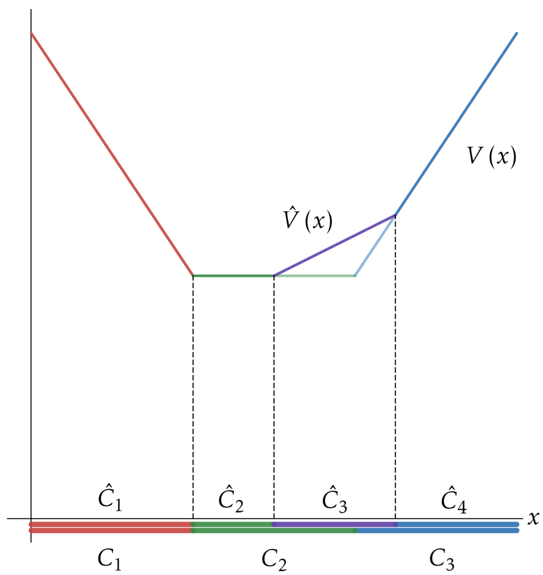
First, we establish that the convexity of $\hat{V} - V$ implies that a transformation generates a greater value for information. This is an implication of a stronger result that is a direct consequence of Blackwell's theorem. For a distribution Φ, $MPC(\Phi)$ denotes the set of mean-preserving contractions (MPCs) of Φ.

Lemma 3.2. *If $\hat{V} - V$ is convex, for any prior $\mu \in \operatorname{int}\Delta(\Theta)$ and distributions over posteriors $\Phi, \Upsilon \in \mathcal{F}_\mu$ with $\Upsilon \in MPC(\Phi)$,*

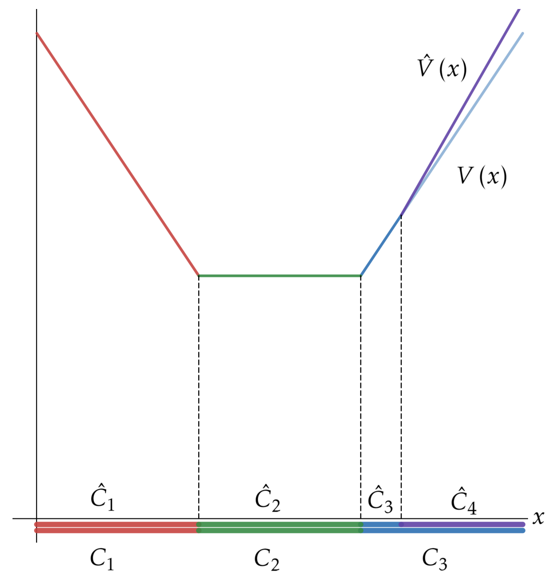
$$\mathbb{E}_\Phi \hat{V}(x) - \mathbb{E}_\Upsilon \hat{V}(x) \geq \mathbb{E}_\Phi V(x) - \mathbb{E}_\Upsilon V(x).$$

Proof. Suppose for the sake of contradiction that

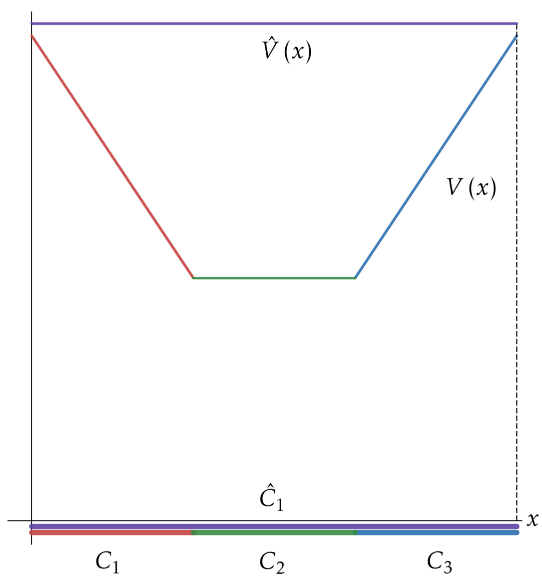
$$\mathbb{E}_\Phi \hat{V}(x) - \mathbb{E}_\Upsilon \hat{V}(x) < \mathbb{E}_\Phi V(x) - \mathbb{E}_\Upsilon V(x),$$



(a) Incomparable C and $\hat{C}$



(b) $\hat{C} \succ C$



(c) $C \succ \hat{C}$

Figure 3: The three possible kinds of new action. There are two states and (initially) three actions. The payoff to the new, fourth, action is depicted in purple. Only in 3b is the new action refining.

which holds if and only if

$$\mathbb{E}_{\Phi}\left[\hat{V}(x) - V(x)\right] < \mathbb{E}_{\Upsilon}\left[\hat{V}(x) - V(x)\right],$$

which violates the definition of an MPC. ■

As no information corresponds to the degenerate distribution on the prior, δ_{μ}, this lemma produces the desired implication, that $\hat{V} - V$ being convex implies that a transformation increases an agent's value for information.

Second, we establish the sufficiency of $\left(\hat{V} - V\right)$'s convexity in the endogenous information case.

Lemma 3.3. *If $\hat{V} - V$ is convex, a transformation does not generate less information acquisition.*

Proof. Fix an arbitrary prior $\mu \in \operatorname{int}\Delta(\Theta)$ and UPS cost functional D. Let Φ^* be an arbitrary solution to Problem $\star$. Suppose for the sake of contradiction that in the transformed decision problem $\hat{\mathcal{D}}$, every optimizer of Problem $\hat{\star}$ is a strict MPC of Φ. Pick one, $\hat{\Phi}^*$. The optimality of $\hat{\Phi}^*$ and strict suboptimality of Φ^* in $\hat{\mathcal{D}}$ imply

$$\mathbb{E}_{\hat{\Phi}^*}\hat{V} - D\left(\hat{\Phi}^*\right) > \mathbb{E}_{\Phi^*}\hat{V} - D\left(\Phi^*\right).$$

Analogously, the optimality of Φ^* in $\mathcal{D}$ implies

$$\mathbb{E}_{\Phi^*}V - D\left(\Phi^*\right) \geq \mathbb{E}_{\hat{\Phi}^*}V - D\left(\hat{\Phi}^*\right).$$

Combining these two inequalities produces

$$\mathbb{E}_{\hat{\Phi}^*}\left[\hat{V} - V\right] > \mathbb{E}_{\Phi^*}\left[\hat{V} - V\right],$$

contradicting that $\hat{\Phi}^*$ is an MPC of Φ^*. ■

Note that in Lemma 3.3 we make no use of the fact that the cost functional is UPS; indeed, we do not even make use of the fact that it is monotone in the Blackwell order. All that is required is that the agent's payoff is additively separable in her value from the decision problem and her cost of acquiring information. Posterior separability of the cost function, instead, disciplines the necessity portion of Theorem 3.1, *infra*. After all, our argument by contraposition would only be made easier by allowing for more general cost functionals (we already know a UPS one will do the trick).

This lemma is not new, though the proof is. The result first appears as Proposition 2 in Chambers, Liu, and Rehbeck (2020); and is, moreover, a corollary of stronger results in Yoder (2022) and Denti (2022). Denti (2022) and Chambers et al. (2020) are similar in aims: both seek to understand which datasets are consistent with costly information acquisition. In his discussion of external validity, Denti (2022) writes "our analysis points to a general property of posterior separable costs that may inform the answer to this question: increasing the incentive to acquire information leads to more extreme beliefs." His ensuing proposition (3) states that if $\hat{V} - V$ is convex, and strictly convex at the prior, then for any solution to the agent's information acquisition problem in the initial decision problem (Problem $\star$), Φ^*, and any solution to the agent's information acquisition problem in the transformed decision problem (Problem $\hat{\star}$), $\hat{\Phi}^*$, the support of $\hat{\Phi}^*$ cannot lie in the relative interior of the convex hull of the support of Φ^*.

A central aspect of Chambers et al. (2020) is their discussion of the implications of additive separability (and whether this property is reasonable) in models of costly information acquisition. To that end, they note that the additively-separable model "forbids an individual from choosing a less informative information structure when there are 'higher gross return from information;' " i.e., $\hat{V} - V$ is convex.[8] In his screening model, in order to characterize the responses to menus of contracts by agents with varying abilities to acquire information, Yoder (2022) establishes a stronger version of Lemma 3.2. His Proposition 4 states that $\left(\hat{V} - V\right)$'s convexity implies that the intersection of the extreme points of any $\hat{\Phi}^*$ with the convex hull of the support of any Φ^* is a (possibly empty) subset of the extreme points of the convex hull of the support of $\hat{\Phi}^*$.

Third, we turn our attention to necessity. It is easiest to start with the endogenous information case.

Lemma 3.4. *If a transformation does not generate less information acquisition, $\hat{V} - V$ is convex.*

[8] Interestingly, both Chambers et al. (2020) and Denti (2022) observe the equivalence of $\left(\hat{V} - V\right)$'s convexity with a transformation generating a greater value for information, but do not give proofs.

Proof. Please visit Appendix A.2. ■

We prove this result by contraposition. If $\hat{V} - V$ is convex, we can construct a cost function that is such that no matter her prior, an agent with value function $\hat{V}$ strictly prefers to acquire no information. In contrast, there are some priors at which an agent with value function V strictly prefers to acquire some information, yielding the result.

Fourth, necessity in the exogenous information case is an easy consequence of the endogenous information lemma. Indeed, we may just take the specific distributions generated in the previous lemma's proof "off-the-shelf."

Lemma 3.5. *If a transformation generates a greater value for information, $\hat{V} - V$ is convex.*

Proof. Contained in Appendix A.3. ■

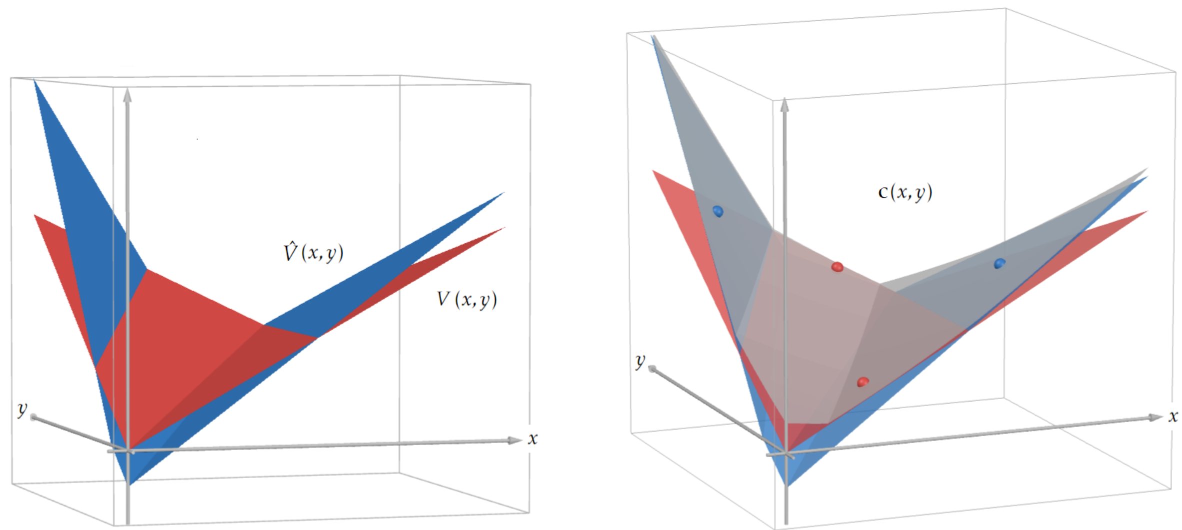
If there are just two states, a stronger statement holds concerning an agent's optimal information acquisition as a result of a transformation. A transformation Generates More Information Acquisition if for any prior $\mu \in \operatorname{int} \Delta(\Theta)$, UPS cost functional D, and solution to the agent's information acquisition problem in the initial decision problem (Problem $\star$), Φ^*, there exists an solution to the agent's information acquisition problem in the transformed decision problem (Problem $\hat{\star}$), $\hat{\Phi}^*$, that is a mean-preserving spread of Φ^*.

Proposition 3.6. *If $|\Theta| = 2$, $\hat{V} - V$ is convex if and only if a transformation generates more information acquisition.*

Proof. Theorem 3.1 implies the necessity portion of the result. Sufficiency is a consequence of the aforementioned Proposition 4 in Yoder (2022). ■

3.1 "More Information" Only in Trivial Cases For Three or More States

When $|\Theta| > 2$, the convexity of $\hat{V} - V$ does not mean that the agent will acquire more information in the transformed decision problem. This is because the new and old experiments may not be (Blackwell) comparable. As we discuss above, Yoder (2022) and Denti (2022) reveal that slightly stronger statements can be made about any solutions to Problems $\star$ and $\hat{\star}$, but these results are still weaker than saying the agent must acquire



(a) V (in red) and $\hat{V}$ (in blue). (b) The desired cost function (in grey).

Figure 4: Proving Proposition 3.8. The distributions supported on the red and blue points are unique solutions to Problems $\star$ and $\hat{\star}$, respectively, yet are Blackwell incomparable.

more information. In this subsection, we show that when there are three or more states, a transformation must lead to more information acquisition only in trivial cases.

For simplicity, we restrict attention to decision problems $\mathcal{D}$ and $\hat{\mathcal{D}}$ containing only a finite number of undominated actions each. We state the following lemma, which may be of independent interest. Let C be the subdivision corresponding to $\mathcal{D}$, and let Φ be a distribution of posteriors whose (finite) support is such that every point $x_i \in \operatorname{supp}\Phi$ lies in the relative interior of a cell $C_i \in C$ and each x_i is in a distinct C_i. We call such an Φ Non-Redundant, as it does not contain multiple posteriors that justify the same action.

We say that a UPS cost D Generates Learning Φ if it is optimal for an agent with cost D to acquire Φ and Φ is uniquely optimal in the sense that any optimally acquired distribution over posteriors is supported on a subset of $\operatorname{supp}\Phi$. Then,

Lemma 3.7. *If Φ is non-redundant, there exists a cost D that generates it.*

Proof. The proof lies in Appendix A.4. ■

Caplin et al. (2022) show that given an UPS cost, for any any Bayes-plausible distribution over posteriors supported on the interior of the simplex, Φ, there is a decision problem that renders Φ optimal in the agent's information-acquisition problem. This lemma, instead, fixes a decision problem, and argues that we can always find a cost function to produce a specified non-redundant Φ.

A potentially interesting implication of this result is toward Bayesian persuasion problems: appealing to the revelation principle, this lemma reveals that a persuader can approximate any optimal Bayesian persuasion outcome even if she cannot directly control the information structure,[9] provided she can specify the agent's flexible cost to acquiring information. That is, a principal can always find a cost function that induces the agent to optimally learn (almost) exactly in the way the principal wants. We illustrate this in the leading prosecutor-judge example of Kamenica and Gentzkow (2011) in Appendix B.

Proposition 3.8. *If $|\Theta| \geq 3$, a transformation generates more information acquisition if and only if $\hat{V} - V$ and/or V is affine.*

Proof. Sufficiency is almost immediate. For necessity, there is a simple proof via contraposition, illustrated in Figure 4. As neither $\hat{V} - V$ nor V is affine, and as decision problems are unaltered by the addition of affine functions, we specify WLOG that there is a region where V lies strictly above $\hat{V}$ (and is not affine on that region), and that the two curves ($\hat{V}$ and V) intersect in such a way that the region where $\hat{V}$ lies strictly above V is not convex. This is illustrated in Figure 4a. Note that we can still do this when there are just two states; in fact, this is one way to deliver a concise proof of Proposition 3.6.

Next, we use Lemma 3.7 to claim the existence of a cost function that intersects $\max\left\{\hat{V}, V\right\}$ at two pairs of points, the line segments between which form a "cross." Furthermore, one of these pairs lies in the region where $\hat{V}$ is strictly larger than V and the the other pair in the region where V is strictly larger than $\hat{V}$. "Cross" refers to the fact that there is a unique intersection point between the line segments connecting each point in a pair; *viz.*, a common point in the (interior of the) convex hull of each pair of points. Note that this also means that the four points do not lie on the same line. This construction

[9] In solutions to Bayesian persuasion problems, there are frequently points of support of optimal distributions that are points of indifference for the agent in her decision problem; and, hence, do not lie in the relative interior of the agent's subdivision's cells.

is illustrated in Figure 4b. This portion of the argument is where three or more states is essential. With just two states, all points are collinear.

By design, when the prior is the meeting point of the two beams of the "cross" and the value function is $\hat{V}$, one of the pairs is uniquely optimal; and when the value function is V, the other is uniquely optimal. As the supports of these binary distributions are not collinear, they are Blackwell incomparable. The details lie in Appendix A.5. ■

4 Four Important Classes of Transformations

4.1 Becoming More Flexible

We begin by transforming an agent's decision problem by adding a single action. We call this Making the Agent More Flexible. In this instantiation, we assume the set of actions initially available to the agent is finite and that in the transformed decision problem, the agent's utility function remains unchanged, $\hat{u} = u$, but her new set of actions is $\hat{A} \coloneqq A \cup \{\hat{a}\}$ for some $\hat{a} \in \mathcal{A} \setminus A$. Our main result of this subsection reveals that subdivisions are central in understanding an agent's comparative value for information.

Theorem 4.1. *A three-way equivalence holds when we make the agent more flexible:*

$\hat{a}$ is refining $\Leftrightarrow$ *A transformation does not generate less information acquisition.* $\Leftrightarrow$ *A transformation generates a greater value for information.*

This theorem is a consequence of Theorem 3.1 and two lemmas. We have already encountered the first lemma, Lemma 2.1, where we establish that $\hat{V} - V$ is convex only if $\hat{C} \succeq C$. Recall that this lemma is not specific to the addition of a single action–it is merely a statement about subdivisions. The second lemma, in contrast, does depend on the fact that it is only a single action that is being added. In §4.2, we elaborate on this point.

Lemma 4.2. *When we make the agent more flexible, $\hat{V} - V$ is convex if $\hat{C} \succeq C$.*

Proof. Please visit Appendix A.6. ■

Theorem 4.1 states that flexibility makes information more valuable for an agent if and only if the new action is weakly dominated or refining. That is, a greater ability to

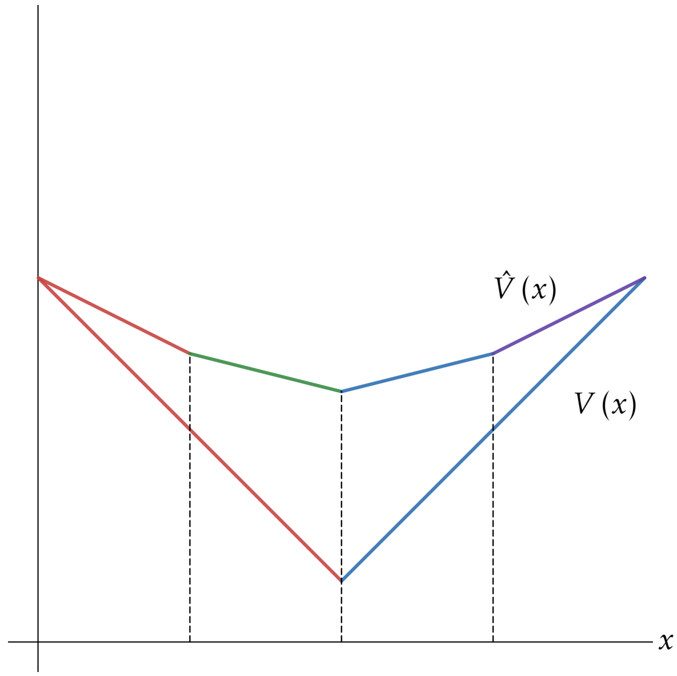


Figure 5: $\hat{C} \succ C$ but $\hat{V} - V$ is not convex.

react to the world increases the value of information if and only if the ability manifests in a particular way. Recall that an action is refining if the regions of beliefs justifying all but one of the previously undominated actions stay unchanged. That non-refining actions may lower the value of information is intuitive: by definition, a non-refining action makes it so that there are now beliefs between which the agent no longer cares to distinguish. Such learning becomes, therefore, worthless.

4.2 Becoming Much More Flexible

Now we transform an agent's decision problem by adding multiple actions. We call this Making the Agent Much More Flexible. We maintain the assumptions that A is finite and that in the transformed decision problem, $\hat{u} = u$. Now, her new set of actions is $\hat{A} := A \cup B$ for some additional finite set of actions $B \in \mathcal{A} \setminus A$.

With multiple new actions, we lose the equivalence of the convexity of $\hat{V} - V$ and $\hat{C}$'s dominance of C in the refinement order. In particular, although Lemma 2.1 states that $\hat{V} - V$ being convex implies $\hat{C} \succeq C$, the converse is false when we make the agent much more flexible. Intuitively, the value function $\hat{V}$ can correspond to a finer subdivision than V but be shallower. In order to preclude such an occurance, we introduce another term. We say that a new set of actions is Totally Refining if each $b \in B$ is either weakly dominated or refining.

Lemma 4.3. *When we make the agent much more flexible, $\hat{V} - V$ is convex if B is totally refining.*

Proof. For any $b \in B$, let V_b denote the agent's value function when the set of actions is $A \cup \{b\}$. Since each b is refining or weakly dominated, $V_b - V$ is convex for all $b \in B$. Finally, $\hat{V} - V = \max(V_b)_{b \in B} - V = \max\{(V_b - V)_{b \in B}\}$ is convex, being the maximum of convex functions. ■

Theorem 3.1 and Lemma 4.3 produce

Corollary 4.4. *Making the agent much more flexible generates a greater value for information and does not generate less information acquisition if the set of additional actions is totally refining.*

The converse to Lemma 4.3 is false, and Figure 6 illustrates this. There, the agent gets access to two new actions, which increases her value of information. Note that the addition of these two actions leaves her subdivision unchanged, i.e., $\hat{C} = C$ but $\hat{V} - V$ is convex.[10] Moreover, the addition of just one of these actions would not increase her value for information. On the other hand, the converse is almost true in the following sense.

We say a new action $\hat{a}$ is Strictly Refining if it is refining and

$$\{x \in \Delta(\Theta) \mid \mathbb{E}_x u(\hat{a}, \theta) \geq V(x)\} \cap C_j = \emptyset,$$

for all $C_j \in C \setminus \{C_i\}$. See Figure 7. A new set of actions is Totally Strictly Refining if each $b \in B$ is either strictly dominated or strictly refining. Given an additional set of actions B, we understand the agent's utility to be an element, u, of the Euclidean space $\mathbb{R}^{B \times \Theta}$ (equipped with the Euclidean metric). Given u, we denote an agent's value function in the transformed decision problem by $\hat{V}^u$. We say that Making the Agent Much More Flexible Generically Generates a Greater Value for Information and Does Not Generate Less Information if $\hat{V}^{\tilde{u}} - V$ is convex for all $\tilde{u}$ in an open ball around u.

[10] I am grateful to Gregorio Curello for suggesting this example.

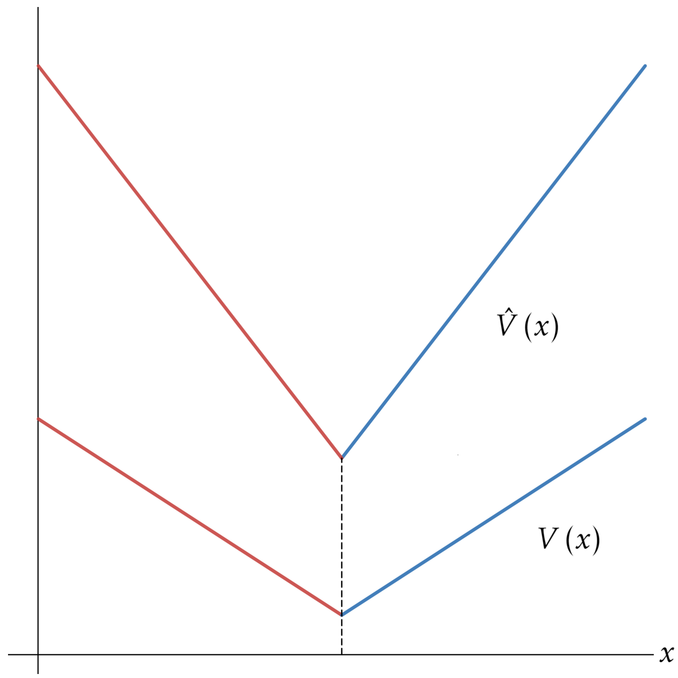
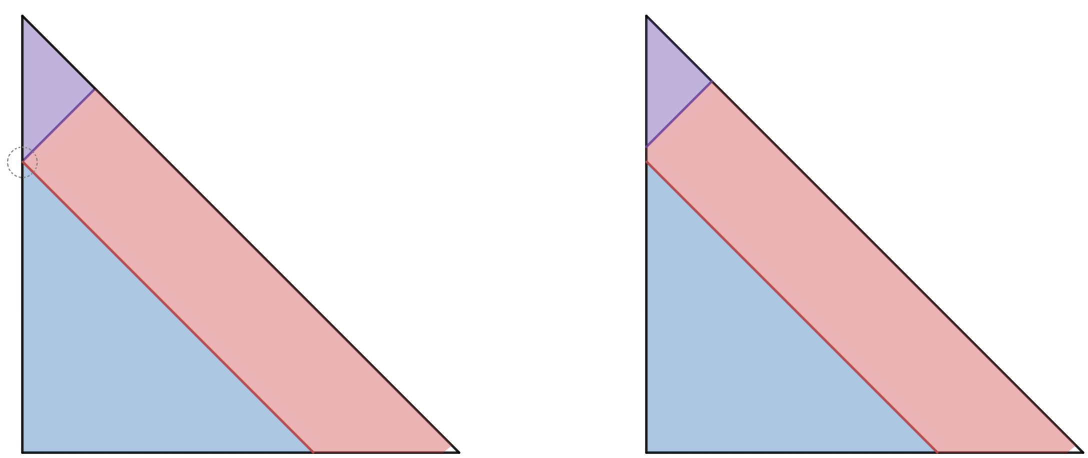


Figure 6: $\hat{C} = C$ and $\hat{V} - V$ is convex. However, B is not totally refining.



(a) Not strictly refining. (b) Strictly refining.

Figure 7: The new action $\hat{a}$ corresponds to the purple cell.

Proposition 4.5. *Making the agent much more flexible generically generates a greater value for information and does not generate less information acquisition if and only if the set of additional actions is totally strictly refining.*

Proof. The proof lies in Appendix A.7. ■

This implies

Corollary 4.6. *Making the agent much more flexible generically generates a greater value for information and does not generate less information acquisition only if the set of additional actions is totally refining.*

4.3 Becoming Much Less Flexible

The generosity of the previous subsections has come to an end. Now, we transform the agent's decision problem by *removing* actions. We call this Making the Agent Much Less Flexible. As in §4.1 and §4.2, A is finite and $\hat{u} = u$. In contrast, in the transformed decision problem, the agent's new set of actions is $\emptyset \neq \hat{A} \subset A$.

If there exists an action $a_i \in \hat{A}$ that is not weakly dominated when the set of actions is A–there exists an $x \in \Delta(\Theta)$ such that $\mathbb{E}_x u(a_i, \theta) > \max_{a \in A \setminus \{a_i\}} \mathbb{E}_x u(a, \theta)$–we say There Are Leftovers. If every eliminated action $a_j \in A \setminus \hat{A}$ is weakly dominated when the set of actions is A–for all $x \in \Delta(\Theta)$, $\mathbb{E}_x u\left(a_j, \theta\right) \leq \max_{a \in A \setminus \{a_j\}} \mathbb{E}_x u(a, \theta)$–we say the elimination is Inconsequential.

Proposition 4.7. *If making the agent much less flexible generates a greater value for information and generates no less information acquisition, there are no leftovers or the removal is inconsequential.*

Proof. The proof lies in Appendix A.8. ■

Clearly, if the pruning of actions is inconsequential, making the agent much less flexible cannot reduce her value for information: her value function is unchanged by the removal of dominated actions. Eliminating actions that are not dominated may yet lead to an increase in her value of information, but only if, in effect, the agent obtains an entirely new decision problem. An increase in her value for information can only occur (when some undominated actions are removed), if *every* undominated action is removed–there can be no leftovers.

Even worse is the situation in which $A \setminus \hat{A}$ is totally refining with respect to $\hat{A}$–*viz.*, starting with set $\hat{A}$, each $a \in A \setminus \hat{A}$ is refining (they are not weakly dominated, by assumption). Then, as Lemma 4.3 implies $V - \hat{V}$ is convex,

Corollary 4.8. *If $A \setminus \hat{A}$ is totally refining with respect to $\hat{A}$, making the agent much less flexible generates a lower value for information and generates no more information acquisition.*

That is

$$\mathbb{E}_{\Phi} \hat{V}(x) - \hat{V}(\mu) \leq \mathbb{E}_{\Phi} V(x) - V(\mu),$$

for all $\Phi \in \mathcal{F}(\mu)$ and $\mu \in \operatorname{int} \Delta(\Theta)$; and for any solution to Problem $\hat{\star}$, $\hat{\Phi}^*$, there exists a solution to Problem $\star$, Φ^*, that is not a strict MPC of $\hat{\Phi}^*$.

It is easiest to understand Proposition 4.7 by reflecting upon the removal of a single action. For information to be more valuable, we know that the new subdivision must be finer than the original one. This means that if there are leftovers and the action we are eliminating is not dominated in A, $\hat{V}$ must equal V on every cell of C but one, C_i. Given this, learning that takes place only within C_i can only increase in value. However, on the relative interior of C_i, $\hat{V}$ must lie strictly below V, which means that for priors outside of C_i, signals that produce posteriors in C_i are more valuable *before* the transformation.

4.4 Transformations of the Agent's Utility Function

Now, the transformation leaves the set of actions available to the agent unchanged; that is, $\hat{A} = A$. Instead, the agent's utility u is modified by some strictly monotone transformation in the transformed decision problem: $\hat{u} = \phi \circ u$ for some strictly monotone ϕ. We call this Transforming the Agent's Utility. A special case is when the transformation is affine. In that instance, $\hat{u} := ku + s$, where $k \in \mathbb{R}_{++}$ and $s \in \mathbb{R}$. Such a transformation preserves the subdivision–$\hat{C} = C$–but how does it affect $\hat{V} - V$?

Proposition 4.9. *An affine transformation of the agent's utility generates a greater value for information and does not generate less information acquisition if and only if* $k \geq 1$.[11]

Proof. The proof resides in Appendix A.9. ■

This affine transformation can be effected in a number of ways. The first is a direct scaling of the payoffs. The second is by repeating the decision problem $k \in \mathbb{N}$ times.

Corollary 4.10. *Repetition of a decision problem (with future payoffs possibly discounted) generates a greater value for information and does not generate less information acquisition.*

A third relates to changes in wealth. The agent's monetary payoff from action a is some function $f_a(\theta)$. Her utility function over terminal wealth w is of the constant absolute risk aversion (CARA) form: $v(w) = -\exp(-\alpha w)$, where $\alpha \in \mathbb{R}_{++}$ is a scaling parameter. Accordingly, the agent's utility as a function of her action and the state θ is

$$u(a, \theta) = -\exp(-\alpha f_a(\theta)) \exp(-\alpha w).$$

Evidently, changing the agent's endowed wealth to $\hat{w}$ produces linear transformation of the utility $u \mapsto ku = \hat{u}$, where

$$k = \frac{\exp(-\alpha \hat{w})}{\exp(-\alpha w)}.$$

Consequently,

Corollary 4.11. *For an agent with CARA utility, decreased wealth generates a greater value for information and does not generate less information acquisition.*

We could also assume that the agent's endowed wealth is random, modeled by a real valued, random variable Y that has a finite mean and is uncorrelated with the state. Y is distributed according to cumulative distribution function H. We say that Aggregate Risk Increases if the distribution of Y changes from H to an MPS of H, $\hat{H}$. As this increase in risk is equivalent to a decrease in wealth for the risk-averse agent,

Corollary 4.12. *For an agent with CARA utility, increased aggregate risk generates a greater value for information and does not generate less information acquisition.*

[11] Denti (2022) also notes this result, but does not give a proof.

One might also wish to say something about how non-affine transformations affect preferences for information. Unfortunately, modifying the agent's utility function by some non-affine transformation may in general change the subdivision;[12] and, in particular, may result in a new subdivision that is incomparable to the previous one. That is, there are some experiments that a more risk-averse agent would value more than a less risk-averse agent, but also experiments for which the reverse holds.[13]

5 Two Applications

5.1 Delegation

In a delegation setting, in which an agent acquires information before taking an action, Szalay (2005) studies how a principal prefers to constrain the agent's set of actions even though their interests are perfectly aligned *ex post*. In particular, Szalay shows that it is optimal for the principal to eliminate "intermediate" actions, which improves incentives for information acquisition. In that spirit, here we note that when the agent chooses whether to buy a fixed experiment, the principal always finds it optimal to increase the agent's flexibility by giving him additional refining actions.

Suppose the principal and agent share the same utility function, a common prior, and that initially the set of actions available to the agent is the finite set A. The agent can acquire information by paying some cost $\gamma > 0$ to see the realization of some signal. After acquiring information, the agent takes an action. The principal can give the agent access to an additional finite set of actions, B, before she acquires information.

Remark 5.1. The principal prefers to give the agent access to an additional set of actions, B, if it is totally refining.

[12] Indeed, as revealed in Weinstein (2016), previously dominated actions may become undominated.

[13] In a related paper, Pease and Whitmeyer (2023), we characterize precisely when the set of beliefs at which one action is preferred to another must increase in size (in a set-inclusion sense)–potentially altering the subdivision–when an agent is made more risk averse.

5.2 Selling Information

Theorem 3.1 suggests a natural analog of increasing differences in an informational setting in which an agent's private type is her value for information. Here we apply this to a monopolistic screening problem. There are n states and an agent has one of two types, ω_1 and ω_2, with respective value functions V_1 and V_2, where $V_1 - V_2$ is convex. The principal and agent share a common prior $\mu \in \operatorname{int}\Delta(\Theta)$ and the principal can "produce" any distribution over posteriors Φ subject to a UPS cost $D(\Phi)$. By the revelation principle, she offers a contract $((t_1, \Phi_1), (t_2, \Phi_2))$.

Naturally, in the first-best problem, the principal solves

$$\max_{\Phi_1 \in \mathcal{F}(\mu)} \left\{ \int_{\Delta(\Theta)} V_1(x)\, d\Phi_1(x) - D(\Phi_1) \right\}, \quad \text{and} \quad \max_{\Phi_2 \in \mathcal{F}(\mu)} \left\{ \int_{\Delta(\Theta)} V_2(x)\, d\Phi_2(x) - D(\Phi_2) \right\},$$

and charges each type a price produced by that type's binding participation constraint. Echoing the basic monopolistic screening model, Theorem 3.1 indicates that in the first-best solution type ω_1 is provided with "no worse quality;" that is, $\Phi_{1,FB}$ is not a strict MPC of $\Phi_{2,FB}$. In addition, Theorem 3.1 tells us that $t_1 \geq t_2$. Naturally, if there are just two states, ω_1 is provided with "higher quality" than type ω_2: $\Phi_{1,FB}$ is an MPS of $\Phi_{2,FB}$

In the second-best problem, following standard logic, IR_2 and IC_1 bind, whereas IR_1 and IC_2 are slack. The principal's objective is therefore (eliminating constants)

$$(1-\rho)\left(\frac{1}{1-\rho}\int_{\Delta(\Theta)} (V_2(x) - \rho V_1(x))\, d\Phi_2(x) - D(\Phi_2)\right) + \rho\left(\int_{\Delta(\Theta)} V_1(x)\, d\Phi_1(x) - D(\Phi_1)\right),$$

where $\rho := \mathbb{P}(\omega_1)$. Since

$$V_2 - \frac{V_2 - \rho V_1}{1-\rho} = \rho \frac{V_1 - V_2}{1-\rho}$$

is convex, Theorem 3.1 tells us that $\Phi_{2,SB}$ cannot be a strict MPC of $\Phi_{2,FB}$ (and for two states, Proposition 3.6 reveals that $\Phi_{2,SB}$ must be an MPC of $\Phi_{2,FB}$). Evidently, $\Phi_{1,SB} = \Phi_{1,FB}$. As we have argued, all of the standard insights go through:

Remark 5.2. In the "selling information" example of this section, in the second-best (screening) solution, there is no output (quality of information) distortion at the top and semi-downward distortion for the "low" type relative to the first-best optimum (strictly more information cannot be provided).

6 Related Work & Discussion

This paper is related to Curello and Sinander (2022), who explore in a single-dimensional setting–either corresponding to a mean-measurable problem with an continuum of states or a binary state–what changes to her indirect payoff lead to greater (or no less) information provision by a persuader. When there are just two states, their first proposition implies Proposition 3.6. The questions they study, as well as those studied here, are fundamentally comparative statics questions, which connects this work to, e.g., Milgrom and Shannon (1994) and Quah and Strulovici (2009). The example of §5.2 is related to Sinander (2022), in which the author exploits his novel converse envelope theorem to show that in a information-sales setting, any Blackwell-increasing information allocation is implementable.

Beyond this, economists (and biologists) have been interested in the value of information for decision makers since Ramsey (1990). This early foray was subsequently followed by the works of Blackwell (Blackwell (1951, 1953)) and Athey and Levin (2018). The list of other related papers studying the value of information includes Donaldson-Matasci, Bergstrom, and Lachmann (2010), who explore the "fitness value of information" from a evolutionary perspective; De Lara and Gossner (2020) who study the value of information using tools from convex analysis; Radner and Stiglitz (1984), De Lara and Gilotte (2007), and Chade and Schlee (2002) who study the marginal valuation of information; and Azrieli and Lehrer (2008), who study preference orders over information structures induced by decision problems.

There are other works related on a technical level. Kleiner, Moldovanu, Strack, and Whitmeyer (2023) study applications of (regular polyhedral) subdivisions to economic settings, with a particular focus on information and mechanism design. Green and Osband (1991) are (to my knowledge) the first to connect decision problems to subdivisions. Lambert (2019) shows that subdivisions are synonymous with elicitability of forecasts. This idea is also present in Frongillo and Kash (2021).

Finally, this paper is also related to the rational inattention literature pioneered by Sims (2003) and furthered by, e.g., Caplin et al. (2022), Chambers et al. (2020), Denti et al. (2022), and Denti (2022). Caplin and Martin (2021) is especially similar in spirit to this paper. There, they formulate a (binary) relation between joint distributions over actions and states: one such joint distribution dominates another if for every utility function, every experiment consistent with the former is more valuable than every experiment consistent with the latter. In this paper, we construct a binary relation between (equivalence classes) of decision problems–one dominates another if information is more valuable in the former. In this spirit, our paper suggests an easy test for Bayesian rationality: give an experimental participant with some initial endowment of bets an additional refining bet; they should be willing to pay more for information as a result.

References

Susan Athey and Jonathan Levin. The value of information in monotone decision problems. *Research in Economics*, 72(1):101–116, 2018.

Yaron Azrieli and Ehud Lehrer. The value of a stochastic information structure. *Games and Economic Behavior*, 63(2):679–693, 2008.

David Blackwell. Comparison of experiments. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 93–102, Berkeley, Calif., 1951. University of California Press.

David Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, 24(2):265–272, 1953.

Andrew Caplin and Daniel Martin. Comparison of decisions under unknown experiments. *Journal of Political Economy*, 129(11):3185–3205, 2021.

Andrew Caplin, Mark Dean, and John Leahy. Rationally inattentive behavior: Characterizing and generalizing shannon entropy. *Journal of Political Economy*, 130(6):1676–1715, 2022.

Hector Chade and Edward Schlee. Another look at the Radner–Stiglitz nonconcavity in the value of information. *Journal of Economic Theory*, 107(2):421–452, 2002.

Christopher P Chambers, Ce Liu, and John Rehbeck. Costly information acquisition. *Journal of Economic Theory*, 186:104979, 2020.

Gregorio Curello and Ludvig Sinander. The comparative statics of persuasion. *Mimeo*, 2022.

Michel De Lara and Laurent Gilotte. A tight sufficient condition for Radner–Stiglitz nonconcavity in the value of information. *Journal of Economic Theory*, 137(1):696–708, 2007.

Michel De Lara and Olivier Gossner. Payoffs-beliefs duality and the value of information. *SIAM Journal on Optimization*, 30(1):464–489, 2020.

Tommaso Denti. Posterior separable cost of information. *American Economic Review*, 112 (10):3215–3259, 2022.

Tommaso Denti, Massimo Marinacci, and Aldo Rustichini. Experimental cost of information. *American Economic Review*, 112(9):3106–23, September 2022.

Matina C. Donaldson-Matasci, Carl T. Bergstrom, and Michael Lachmann. The fitness value of information. *Oikos*, 119(2):219–230, 2010.

Rafael M Frongillo and Ian A Kash. General truthfulness characterizations via convex analysis. *Games and Economic Behavior*, 130:636–662, 2021.

Edward J. Green and Kent Osband. A revealed preference theory for expected utility. *The Review of Economic Studies*, 58(4):677–695, 06 1991.

Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *The American Economic Review*, 101(6):2590–2615, 2011.

Andreas Kleiner, Benny Moldovanu, Philipp Strack, and Mark Whitmeyer. Applications of power diagrams to mechanism and information design. *Mimeo*, 2023.

Nicholas S. Lambert. Elicitation and evaluation of statistical forecasts. *Mimeo*, 06 2019.

Carl W Lee and Francisco Santos. 16: Subdivisions and triangulations of polytopes. *Handbook of Discrete and Computational Geometry. Ed. by Jacob E. Goodman, Joseph O'Rourke, and Csaba D. Tóth*, pages 415–447, 2017.

Paul Milgrom and Chris Shannon. Monotone comparative statics. *Econometrica*, 62(1): 157–180, 1994.

Marilyn Pease and Mark Whitmeyer. Safety, in numbers. *Mimeo*, Oct 2023.

John W Pratt. Risk aversion in the small and in the large. *Econometrica*, 32(1/2):122–136, 1964.

John K.-H. Quah and Bruno Strulovici. Comparative statics, informativeness, and the interval dominance order. *Econometrica*, 77(6):1949–1992, 2009.

Roy Radner and Joseph Stiglitz. A nonconcavity in the value of information. In *M. Boyer, R. Kihlstrom (Eds.), Bayesian Models of Economic Theory*, pages 33–52. Elsevier, Amsterdam, 1984.

F. P. Ramsey. Weight or the value of knowledge. *The British Journal for the Philosophy of Science*, 41(1):1–4, 1990.

Ron Siegel and Bruno Strulovici. The economic case for probability-based sentencing. *Mimeo*, 2020.

Christopher A Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690, 2003.

Ludvig Sinander. The converse envelope theorem. *Econometrica*, 90(6):2795–2819, 2022.

Dezsö Szalay. The economics of clear advice and extreme options. *The Review of Economic Studies*, 72(4):1173–1198, 10 2005.

Jonathan Weinstein. The effect of changes in risk attitude on strategic behavior. *Econometrica*, 84(5):1881–1902, 2016.

Nathan Yoder. Designing incentives for heterogeneous researchers. *Journal of Political Economy*, 130(8):2018–2054, 2022.

A Omitted Proofs

A.1 Lemma 2.1

Proof. Suppose for the sake of contraposition $\hat{C} \not\succeq C$. This implies for some $\hat{C}_j \in \hat{C}$, there is no C_i such that $\hat{C}_j \succeq C_i$. This means there exist $x, x' \in \hat{C}_j$ and some $C_i \in C$ with $x \in C_i$ and $x' \notin C_i$. By definition $\hat{V}$ is affine on $\hat{C}_j$, so for all $\lambda \in [0,1]$,

$$\lambda \hat{V}(x) + (1-\lambda)\hat{V}(x') = \hat{V}(\lambda + (1-\lambda)x').$$

By construction, for all $\lambda \in (0,1)$,

$$\lambda V(x) + (1-\lambda)V(x') > V(\lambda + (1-\lambda)x').$$

Combining these, and maintaining the convention $\hat{W} = \hat{V} - V$, we obtain that

$$\lambda \hat{W}(x) + (1-\lambda)\hat{W}(x') < \hat{W}(\lambda + (1-\lambda)x'),$$

so $\hat{V} - V$ is not convex. ■

A.2 Lemma 3.4 Proof

Proof. Suppose for the sake of contraposition that $\hat{V} - V$ is not convex on $\Delta(\Theta)$. As $\hat{V} - V$, being the difference of two continuous functions, is continuous, it is not convex on $\operatorname{int}\Delta(\Theta)$.

Let $\rho(x)$ be some strictly convex function on $\Delta(\Theta)$ and for an arbitrary $\varepsilon > 0$ define function $c_\varepsilon(x) := \varepsilon\rho(x) + \hat{V}(x)$. By construction, for all $\varepsilon > 0$, c_ε is strictly convex.

Moreover, $\hat{V} - c_\varepsilon = -\varepsilon\rho$ is strictly concave for all $\varepsilon > 0$, so in the agent's flexible information acquisition problem for the transformed decision problem, the unique solution is for her to acquire the degenerate distribution on her prior, δ_μ, for any prior $\mu \in \operatorname{int}\Delta(\Theta)$.

In contrast, in the initial decision problem, $\mathcal{D}$, the agent's objective in her flexible information acquisition problem is

$$V - c_{\varepsilon} = V - \hat{V} - \varepsilon\rho.$$

As $W := V - \hat{V}$ is not concave on $\operatorname{int}\Delta(\Theta)$, for all sufficiently small $\varepsilon > 0$,

$$\lambda(W(x) - \varepsilon\rho(x)) + (1-\lambda)(W(x') - \varepsilon\rho(x')) > W(\lambda x + (1-\lambda)x') - \varepsilon\rho(\lambda x + (1-\lambda)x') \quad (1)$$

for some $\lambda \in (0,1)$ and $x, x' \in \operatorname{int}\Delta(\Theta)$. Accordingly, as we may set $\mu = \lambda x + (1-\lambda)x'$, there exists a $\mu \in \operatorname{int}\Delta\Theta$ such that the agent strictly prefers acquiring some information to learning nothing. That is, for any optimal Φ^*, $\hat{\Phi}^* = \delta_{\mu}$ is a strict MPC of Φ^*. ■

A.3 Lemma 3.5 Proof

Proof. Suppose again for the sake of contraposition that $\hat{V} - V$ is not convex, without loss of generality on $\operatorname{int}\Delta(\Theta)$. Inequality 1 implies there exist $\lambda \in (0,1)$, $x, x' \in \operatorname{int}\Delta(\Theta)$, and $\mu = \lambda x + (1-\lambda)x'$ such that

$$\lambda\left(V(x) - \hat{V}(x)\right) + (1-\lambda)\left(V(x') - \hat{V}(x')\right) > V(\mu) - \hat{V}(\mu),$$

which holds if and only if

$$\mathbb{E}_{\Phi} V(x) - V(\mu) > \mathbb{E}_{\Phi}\hat{V}(x) - \hat{V}(\mu),$$

where Φ is the binary Bayes-plausible distribution with support $\{x, x'\}$. ■

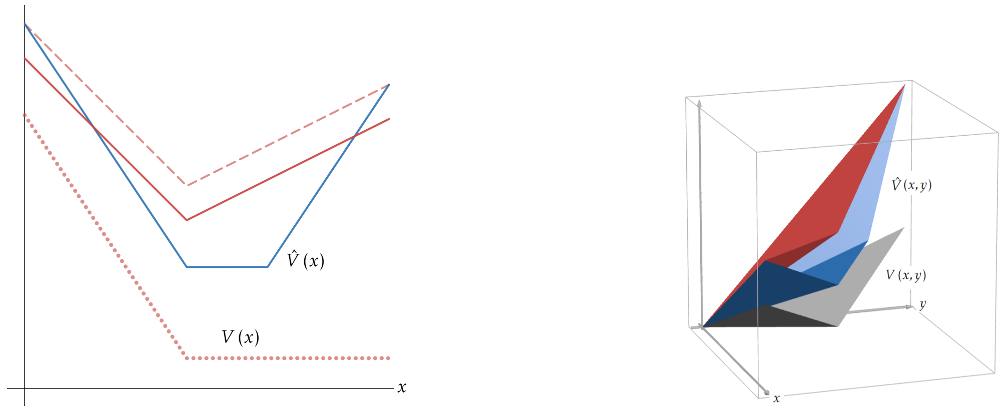
A.4 Lemma 3.7 Proof

Proof. As there are only finitely many undominated actions,

$$V(x) = \max_{i \in I}\{\alpha_i \cdot x + \beta_i\},$$

where $\alpha_i, \beta_i \in \mathbb{R}$ for all $i \in I := \{1, \ldots, m\}$. Let $(x_1, \cdots, x_s) \equiv \operatorname{supp}\Phi$ for a non-redundant Φ $(1 \le s \le m)$. We define

$$t_i(x) = \begin{cases} \alpha_i \cdot x + \beta_i + \varepsilon\|x - x_i\|^2, & \text{if} \quad x_i \in \operatorname{supp}\Phi \\ \alpha_i \cdot x + \beta_i + \varepsilon\|x - x_1\|^2, & \text{if} \quad x_i \notin \operatorname{supp}\Phi, \end{cases}$$



(a) V (red dots), $V+H$ (red dashed), $V+H-\varepsilon$ (solid red) and $\hat{V}$ (solid blue). Use the slider to vary ε.

(b) V (grey), $V+H$ (red), and $\hat{V}$ (blue). Use the slider to vary ε.

Figure 8: Justifying the specification in Proposition 3.8.

where $\varepsilon > 0$.

Then, letting $c(x) = \max_{i \in I} t_i(x)$ it is easy to see that c is strictly convex; $V(x) - c(x) \leq 0$ for all $x \in \Delta(\Theta)$; and for all sufficiently small ε, $V(x) - c(x) = 0$ if and only if $x \in \operatorname{supp}\Phi$. The result follows. ■

A.5 Proposition 3.8 Proof

Proof. ($\Leftarrow$) If V is affine, $V - c$ is strictly concave, so it is uniquely optimal to acquire no information in Problem ⋆. Consequently, any solution to Problem ⋆ is an MPC of the solution to Problem $\hat{\star}$. If $\hat{V} - V$ is affine, the solutions to problems ⋆ and $\hat{\star}$ are identical.

($\Rightarrow$) We prove the result by contraposition. Suppose neither V nor $\hat{V} - V$ is affine. If $\hat{V} - V$ is not convex, by Theorem 3.1, we are done, so let $\hat{V} - V =: \hat{W}$ be convex. As neither $\hat{W}$ nor V is affine and as we can add affine functions to value functions without altering decision problems, we specify WLOG that the sets $Z := \left\{x \in \Delta(\Theta) : \hat{W}(x) \leq 0\right\}$ and $Y := \left\{x \in \Delta(\Theta) : \hat{W}(x) > 0\right\}$ are such that

(i) Z is a polytope of full dimension in $\mathbb{R}^{n-1}$;

(ii) Neither V nor $\hat{W}$ is affine on Z

(iii) Y is of full dimension in $\mathbb{R}^{n-1}$; and

(iv) There exist two points $x_1, x_2 \in \operatorname{int} Y$ and $\lambda \in (0,1)$ for which $\mu = \lambda x_1 + (1-\lambda) x_2 \in \operatorname{int} Z$ and $\hat{V}$ is locally affine around both x_1 and x_2.[14]

Define $\tilde{W}(x) := \max\left\{V(x), \hat{V}(x)\right\}$ and let $\tilde{C}$ be the corresponding subdivision. By construction, there are at least two cells $\tilde{C}_1, \tilde{C}_2 \subseteq Z$. Take x_1, x_2 and μ as previously specified. By construction, we can find $\tilde{x}_1 \in \operatorname{int} \tilde{C}_1$, $\tilde{x}_2 \in \operatorname{int} \tilde{C}_2$ and $\tilde{\lambda} \in (0,1)$ such that $\mu = \tilde{\lambda} \tilde{x}_1 + \left(1 - \tilde{\lambda}\right) \tilde{x}_2$ and $\tilde{x}_1$ and $\tilde{x}_2$ do not lie on the line segment between x_1 and x_2.

Let Φ be a distribution with support $\{x_1, x_2, \tilde{x}_1, \tilde{x}_2\}$. By construction, Φ is non-redundant when the agent's value function is $\tilde{W}$. Consequently, Lemma 3.7 implies there is a cost function, D, that generates Φ. Moreover, for prior μ, D renders the Bayes-plausible binary distribution with support $\{x_1, x_2\}$ uniquely optimal when the agent's value function is $\hat{V}$ and the Bayes-plausible binary distribution with support $\{\tilde{x}_1, \tilde{x}_2\}$ uniquely optimal when the agent's value function is V. These two distributions are Blackwell-incomparable. ■

A.6 Lemma 4.2 Proof

Proof. Suppose $\hat{C} \geq C$. That is, for every $\hat{C}_j \in \hat{C}$, there exists a $C_i \in C$ such that $\hat{C}_j \subseteq C_i$. Define

$$\hat{W}(x) := \hat{V}(x) - V(x) = \max\left\{\hat{\alpha} \cdot x + \hat{\beta}, V(x)\right\} - V(x),$$

where $\hat{\alpha} \cdot x + \hat{\beta}$ ($\hat{\beta} \in \mathbb{R}$, $\hat{\alpha} \in \mathbb{R}^{n-1}$) is the expected payoff from taking the new action $\hat{a}$. By assumption,

$$\hat{C}_j := \left\{x \in \Delta \colon \hat{\alpha} \cdot x + \hat{\beta} \geq V(x)\right\} \subseteq C_j$$

for some $j \in \{1, \ldots, m\}$. Moreover, by the continuity of V and $\hat{V}$, for all $x \in \cup_{i \neq j} C_i$, $\hat{W}(x) = V(x) - V(x) = 0$.

For all $x \in C_j$, $V(x) = \alpha \cdot x + \beta$ ($\beta \in \mathbb{R}$, $\alpha \in \mathbb{R}^{n-1}$). Accordingly, for all $x \in C_j$,

$$\hat{W}(x) = \max\left\{\hat{\alpha} \cdot x + \hat{\beta}, \alpha \cdot x + \beta\right\} - (\alpha \cdot x + \beta) = \max\left\{0, (\hat{\alpha} - \alpha) \cdot x + \hat{\beta} - \beta\right\},$$

[14]To see that this specification is WLOG, take arbitrary non-affine V and $\hat{V}$, where $\hat{V} - V$ is convex and not affine. Evidently, there exists a hyperplane $H(x) = \alpha \cdot x + \beta$ such that $V + H = \hat{V}$ at each vertex of the simplex. As $\hat{V} - V$ is convex and not affine, $V + H$ lies weakly above $\hat{V}$ on $\Delta(\Theta)$ and strictly above $\hat{V}$ on $\operatorname{int} \Delta\Theta$. Then, just subtract a sufficiently small constant ε from $V + H$. Figure 8 depicts this in 2d and 3d.

which means that for all $x \in \Delta(\Theta)$,

$$\hat{W}(x) = \max\left\{0, (\hat{\alpha} - \alpha) \cdot x + \hat{\beta} - \beta\right\},$$

which is convex, being the maximum of two affine functions. ■

A.7 Proposition 4.5 Proof

Proof. (⇒) This direction mostly repeats the proof of Lemma 4.3. Fix an arbitrary $b \in B$ and let $\hat{V}_b^u$ denote the agent's value function when the set of actions is $A \cup \{b\}$ and the utilities from the new actions are $u \in \mathbb{R}^{B \times \Theta}$. Since b is either strictly refining or strictly dominated, $\hat{V}_b^u - V$ is convex. Moreover, if b is strictly dominated, $\mathbb{E}_x u(b,\theta) < V(x)$ for all $x \in \Delta(\Theta)$. Consequently, there exists an open ball around u such that for all $\tilde{u}$ in that open ball, $\mathbb{E}_x u(b,\theta) < V(x)$ for all $x \in \Delta(\Theta)$. Now let b be strictly refining. Let C_i be the cell of C for which the following inclusion holds:

$$\{x \in \Delta(\Theta) \mid \mathbb{E}_x u(\hat{a},\theta) \geq V(x)\} \subseteq C_i.$$

By the definition of strictly refining, for all $x \notin C_i$, $\mathbb{E}_x u(b,\theta) < V(x)$. Accordingly, there exists an open ball around u such that for all $\tilde{u}$ in that open ball, $\mathbb{E}_x u(b,\theta) < V(x)$ for all $x \in \Delta(\Theta) \setminus C_i$. Combining these two observations, we see that for all $\tilde{u}$ in some open ball around u, b is strictly refining, so $\hat{V}_b^{\tilde{u}} - V$ is convex. Finally, $\hat{V}^{\tilde{u}} - V = \max\left(V_b^{\tilde{u}}\right)_{b \in B} - V = \max\left\{\left(V_b^{\tilde{u}} - V\right)_{b \in B}\right\}$ is convex, being the maximum of convex functions.

(⇐) Suppose for the sake of contraposition that B is not totally strictly refining. That is, there exists a $b \in B$, an $x \in \Delta$, and two undominated (in A) actions $a_1, a_2 \in A$ for which

$$\mathbb{E}_x u(b,\theta) \geq \mathbb{E}_x u(a_1,\theta) = \mathbb{E}_x u(a_2,\theta) = V(x).$$

Without loss of generality, we may assume $\hat{V}(x) = \mathbb{E}_x u(b,\theta)$ as any $b' \neq b$ for which $\mathbb{E}_x u(b',\theta) = \hat{V}(x) > \mathbb{E}_x u(b,\theta)$ is neither dominated nor strictly refining itself, so we could just replace b in the proof with b'.

Now, pick some $\theta' \in \Theta$ that occurs with positive probability under x. Define $\tilde{u}(b,\theta) = u(b,\theta)$ for all $\theta \neq \theta'$ and $\tilde{u}(b,\theta') = u(b,\theta') + \varepsilon$. Then, for all $\varepsilon > 0$,

$$\mathbb{E}_x \tilde{u}(b,\theta) > \hat{V}(x) \geq \mathbb{E}_x u(a_1,\theta) = \mathbb{E}_x u(a_2,\theta) = V(x). \tag{2}$$

By the continuity of each of the four functions, $\mathbb{E}_x \tilde{u}(b,\theta)$, $\hat{V}(x)$, $\mathbb{E}_x u(a_1,\theta)$, and $\mathbb{E}_x u(a_2,\theta)$ in x, for all x' in some open ball (understanding $\Delta(\Theta)$ as a subset of $\mathbb{R}^{n-1}$, equipped with the Euclidean metric) around x,

$$\mathbb{E}_{x'} \tilde{u}(b,\theta) > \max\left\{\hat{V}(x'), \mathbb{E}_{x'} u(a_1,\theta), \mathbb{E}_{x'} u(a_2,\theta), V(x')\right\}.$$

Moreover, neither a_1 nor a_2 is weakly dominated so for any open ball $B_\eta(x)$ around x, there exist $x_1, x_2 \in B_\eta(x)$ such that $\mathbb{E}_{x_1} u(a_1,\theta) > \max_{a\in A\setminus\{a_1\}} \mathbb{E}_{x_1} u(a,\theta)$ and $\mathbb{E}_{x_2} u(a_2,\theta) > \max_{a\in A\setminus\{a_2\}} \mathbb{E}_{x_2} u(a,\theta)$.

This implies there is no $C_i \in C$ such that

$$\left\{x \in \Delta(\Theta) \mid \mathbb{E}_x \tilde{u}(b,\theta) \geq \max_{a\in A\cup B} \mathbb{E}_x u(a,\theta)\right\} \subseteq C_i,$$

so $C \not\succeq \hat{C}^{\tilde{u}}$, where $\hat{C}^{\tilde{u}}$ denotes the subdivision in the perturbed transformed decision problem. The contraposition of Lemma 2.1, therefore, produces the result. ■

A.8 Proposition 4.7 Proof

Proof. Suppose for the sake of contraposition there are leftovers and the elimination is not inconsequential. Namely, there is some $a_i \in \hat{A}$ for which there exists an $x \in \Delta(\Theta)$ such that

$$\mathbb{E}_x u(a_i,\theta) > \max_{a\in A\setminus\{a_i\}} \mathbb{E}_x u(a,\theta);$$

and there is some $a_j \in A \setminus \hat{A}$ for which there exists an $x \in \Delta(\Theta)$ such that

$$\mathbb{E}_x u\left(a_j,\theta\right) > \max_{a\in A\setminus\left\{a_j\right\}} \mathbb{E}_x u(a,\theta).$$

By construction, for all $x \in C_i$, $\mathbb{E}_x u(a_i,\theta) = V(x) = \hat{V}(x)$ and for all $x \in \operatorname{int} C_i$,

$$\mathbb{E}_x u(a_i,\theta) = V(x) = \hat{V}(x) > \max_{a\in A\setminus\{a_i\}} \mathbb{E}_x u(a,\theta) \geq \max_{a\in \hat{A}\setminus\{a_i\}} \mathbb{E}_x u(a,\theta).$$

Let $a_j \in A \setminus \hat{A}$; namely, a_j is one of the actions taken away. For all $x \in \operatorname{int} C_j$,

$$\mathbb{E}_x u\left(a_j,\theta\right) = V(x) > \hat{V}(x).$$

By construction, there exist points $x', \mu \in \operatorname{int} C_i$, $x \in \operatorname{int} C_j$, and weight $\lambda \in (0,1)$ such that $\mu = \lambda x + (1-\lambda) x'$, and

$$\lambda \underbrace{\left(\hat{V}(x) - V(x)\right)}_{<0} + (1-\lambda) \underbrace{\left(\hat{V}(x') - V(x')\right)}_{=0} - \underbrace{\left(\hat{V}(\mu) - V(\mu)\right)}_{=0} < 0,$$

so $\hat{V} - V$ is not convex. ■

A.9 Proposition 4.9 Proof

Proof. If $k = 1$, the result is immediate, so let $k \neq 1$. For any pair of points $x_1 \neq x_2 \in \Delta(\Theta)$ with corresponding optimal actions $a_1 \neq a_2$, $\lambda \in [0,1]$, and optimal action at $x^\dagger := \lambda x_1 + (1-\lambda) x_2$, $a^\dagger$,

$$\begin{aligned} \lambda \hat{W}(x_1) + (1-\lambda) \hat{W}(x_2) > \hat{W}\left(x^\dagger\right) \quad &\Leftrightarrow \\ (k-1)\left[\lambda \mathbb{E}_{x_1} u(a_1, \theta) + (1-\lambda) \mathbb{E}_{x_2} u(a_2, \theta)\right] > (k-1) \mathbb{E}_{x^\dagger} u\left(a^\dagger, \theta\right) \quad &\Leftrightarrow \\ k > 1, \end{aligned}$$

recalling that $\hat{W}(x) := \hat{V}(x) - V(x)$. ■

B Using Lemma 3.7 in Bayesian Persuasion

Here, we work through the leading prosecutor-judge example of Kamenica and Gentzkow (2011). The receiver's value function is $V(x) = \max\{x, 1-x\}$, the prior is $\frac{3}{10}$ and the sender wishes to maximize the probability that the receiver takes the "high action," which the receiver does if and only if her posterior is above $\frac{1}{2}$. The unique persuasion solution is the binary distribution with support $\left\{0, \frac{1}{2}\right\}$. By Lemma 3.7, we can construct a cost function that approximates it:

Remark B.1. For any $\eta \in \left(0, \frac{1}{2}\right)$, there exists a UPS cost that induces the receiver to optimally acquire the binary distribution with support $\left\{\eta, \frac{1}{2} + \eta\right)$.

Proof. Following Lemma 3.7, here is a cost function that does the trick:

$$c(x) = \max\left\{x + \varepsilon\left(x - \frac{1}{2} - \eta\right)^2, 1 - x + \varepsilon (x - \eta)^2\right\},$$

for any $\varepsilon \in (4 - 8\eta, 8\eta)$. Here is an interactive graph of the solution. ■