Temptation and Persuasion

Alexander M. Jakobsen*

March 11, 2025

Abstract

I analyze the foundations, comparative statics and identification properties of Bayesian Persuasion in terms of the receiver's choices, preferences and welfare. By interpreting sender and receiver as different selves, the model delivers new insights into the interaction between temptation, attention, demand for commitment, and the value of information. All model parameters—and, thereby, endogenous signals—are identified from receiver's behavior, enabling simple comparative statics regarding the degree of conflict between agents and measures of optimism or pessimism. My results differ substantially from those of related decision models, highlighting the role of *motivated attention* in intrapersonal conflict.

1 Introduction

1.1 Motivation

The Bayesian Persuasion model (Kamenica and Gentzkow, 2011) has received considerable attention and become a central framework in the economics of information design and disclosure. In the standard setup, one agent (Sender) selects an information structure and commits to revealing its signal realization to another (Receiver). The literature has examined numerous extensions and variations of the baseline model, with results typically focusing on Sender's choice of information and whether he benefits from this opportunity. This paper examines the interaction from Receiver's perspective. In particular, I develop a decision-theoretic analogue of Bayesian Persuasion that takes as primitive Receiver's preferences or choices; Sender's behavior is not directly observed and must be inferred from Receiver's.

*Kellogg MEDS, Northwestern. I am grateful to Nemanja Antic, Josephine Auer, Ian Ball, Tim Feddersen, Carlos Gonzalez Perez, Peter Klibanoff, Jay Lu, Philipp Sadowski, Ron Siegel, Vasiliki Skreta, Max Stinchcombe, Ina Taneva, Ben Young and audience members at Arizona, Columbia, Duke/UNC, the Hitotsubashi Summer Institute 2022, Michigan, Northwestern, NYU, Penn State, SAET 2023, Toronto, UCLA, UPenn, and UT Austin for helpful feedback. This paper was previously entitled "Revealed Persuasion."

My motivation is twofold. First, the decision-theoretic approach allows several basic questions about the identification and comparative statics of persuasion to be resolved at a high level of generality. For example, when might Receiver benefit from commitment (shrinking the action set) or from additional public information? Each avenue serves to level the playing field against Sender's commitment power but distorts incentives in nontrivial ways. Nonetheless, these, and other, questions resolve quite naturally in my framework. Identification is also of central importance: does Receiver's behavior reveal parameter values (utility functions and prior beliefs), or is richer data needed? It turns out that various types of Receiver data are sufficient to identify all parameters—hence unobserved information provision—and my results provide methods to perform such elicitation.

Second, the Receiver-based model of persuasion, interpreted as an *intrapersonal* interaction, provides a theory of behavior delivering new insights into widely-studied phenomena such as temptation, (in)attention, and demand for commitment. This approach—similar in spirit to self-signaling models in behavioral economics—yields new implications for, and links between, the value of commitment and of information for agents facing temptation or other competing motives when making decisions.

To illustrate, consider an individual deciding between two options at a restaurant: a burger or a salad.[1] The burger is always the same but the salad varies with exogenous factors like the availability of ingredients, how busy the kitchen is, and so on; in effect, there are two types of salad and an exogenous state determines which type will arrive if the individual orders one. The individual prefers the first type of salad to the burger and the burger to the second type, and prior beliefs are such that the salad is optimal ex ante. These preferences are represented by a utility index u and prior beliefs μ. "Temptation" preferences involve the same prior but a different utility index v; in this case, v ranks the burger strictly preferred to either type of salad. What will the individual do?

The literature has provided a variety of answers. Strotzian models (Strotz, 1955), for example, involve changing utility functions: the ex-ante function u changes to v at the time of consumption. Thus, the burger is chosen with probability 1. In the Gul and Pesendorfer (2001) model, utilities do not change over time. Instead, the individual maximizes u-utility minus the "cost of self-control": the gap in v-utility between the chosen option and the v-maximal option. Thus, temptation-utility v influences, but does not dictate, consumption choice: depending on u and v, the individual may choose the burger or exercise costly self-control and choose the salad. Many models building on Strotz (1955), Gul and Pesendorfer (2001), and others have been developed.

The mechanism in this paper is fundamentally different. Like the Gul and Pesendorfer

[1] This is a slightly-modified version of the leading example in Gul and Pesendorfer (2001).

(2001) model, my approach combines temptation with self-control. However, the form of self-control and the manner in which v influences choice are different. Specifically, I treat u as Receiver's utility function in Bayesian Persuasion and v as Sender's. This means information is chosen to maximize expected v-utility given that signal-contingent consumption choices maximize expected u. Thus, temptation compels the individual to pay *selective attention* in order to increase the probability of the burger being chosen; u and μ are such that the salad is preferred ex-ante but selective (or *motivated*) attention can generate acceptable *justifications* (signal realizations) for choosing the burger. That choices must be justified is precisely how the individual exercises self-control, and selective attention provides a way to increase temptation-utility v subject to that constraint.[2]

In general, the intrapersonal (or "behavioral") interpretation of the model captures situations where the same individual or entity (i) acquires information and subsequently makes choices, (ii) experiences temptation or other competing motives, and (iii) must justify choices, using available information, as consistent with their "true" objective.

Example 1. Consider the following scenarios:

(a) A voter values candidates according to their policy positions (utility u) but may be swayed by other characteristics like charisma or party affiliation (utility v). Selective attention to news, debate highlights, etc can provide policy-based rationales for supporting candidates of greater v-utility.

(b) An investor decides which firm(s) to invest in; u weighs both profitability and social responsibility while v represents "selfish" preferences that only consider the investor's returns. Selective attention to financial statements, environmental impact reports, etc can skew u-optimal investment decisions in favor of more profitable firms.

(c) An organization conducts a study to determine whether to fund a project; u reflects the publicly-stated objectives of the organization and v those of its members. Carefully-crafted studies can advance priorities v while justifying choices as consistent with u. ⧫

These scenarios highlight the role of motivated attention in resolving conflict between objectives u and v. They also suggest why choices must be justified as u-optimal. In (a) and the

[2]In keeping with the standard Bayesian Persuasion framework, there is no cost of acquiring information. For example, the individual might ask the server what kind of salad is available (perfect information) or opt for a noisier signal (peripheral glances at other tables or the kitchen line, recent online reviews, etc). To some extent, bounds on information are implicit in the (exogenous) state space. A more elaborate model might explicitly introduce costly information but here I deliberately focus on how intrapersonal conflict affects incentives for acquiring freely-available information.

restaurant example above, psychological constraints like willpower or a desire to maintain a positive self-image can create a need to justify choices; similarly, self-image or ethical considerations might lead to the behavior in (b). There are many precedents for such self-imposed constraints in the literature.[3] In contrast, the constraint in (c) is imposed externally by the public, who demand that the organization uphold standards u.[4]

In general, information acquisition serves to "reverse engineer" justifications for tempting actions. The agent sacrifices some u-utility by garbling free information but is otherwise satisfied with the plausible deniability offered by signal realizations; acceptance of such tailored information despite the availability of perfect information is what distinguishes this from standard rational behavior.[5] Note, however, that since the distribution of posteriors averages out to the prior, individuals in my model cannot systematically bias themselves.[6]

As alluded to above, this mechanism for resolving intrapersonal conflict incorporates elements of the influential self-signaling paradigm (Bénabou and Tirole, 2002) in behavioral economics. Despite overlapping motivations, this strand of literature has evolved separately from the decision-theoretic approach to intrapersonal conflict. This paper does not unify these strands but, by bringing elements of self-signaling to a decision-theoretic setting, provides key insights into its foundations, comparative statics and identification properties. My approach also facilitates comparisons to established axiomatic models of intrapersonal conflict; as we shall see, motivated attention has rather different implications for behavior.

1.2 Results

A single decision maker, Receiver, chooses among *acts* (Anscombe and Aumann, 1963): profiles $f = (f_\omega)_{\omega \in \Omega}$ of lotteries $f_\omega \in \Delta X$ where Ω is a finite set of states, X a finite

[3]For example, there is an active literature on justifiable choice (see section 1.3). Moral considerations or self-image concerns are among several reasons why individuals feel a need to justify choices. Adam Smith (Smith, 1759) argues that "we endeavour to examine our own conduct as we imagine any other fair and impartial spectator would examine it." Here, one might interpret u as an individual's desired "type"; in order to maintain a positive (perhaps moral) self-image, the individual requires choices to be consistent with u. Such consistency resembles a *personal rule* or, as Ainslie (1992) describes it, "the kind of impulse control ... which allows a person to resist impulses while he is both attracted by them and able to pursue them." Bénabou and Tirole (2004) develop a model capturing these ideas.

[4]Similarly, individuals may feel compelled to signal their "type" u to others. Such image (or social signaling) concerns are especially prevalent in settings of charitable giving and related activity; see, for example, Bénabou and Tirole (2006), Grossman (2015), and Exley and Kessler (2023).

[5]This is similar to the concept of "moral wiggle room" (Dana et al., 2007) in contexts of other-regarding behavior (and within the scope of my model), where uncertainty about the consequences of actions "justifies" selfishness. This incentivizes various forms of willful ignorance (Haisley and Weber, 2010; Exley, 2016; Grossman and Van Der Weele, 2017; Robbett et al., 2024) or "motivated errors" (Exley and Kessler, 2024).

[6]That is, agents in my model need not choose freely-available perfect information but are otherwise Bayesian in their analysis of acquired information. Many other theories of information avoidance (belief-based utility, optimism maintenance, etc) do not have this property; see Golman et al. (2017) for a survey.

set of outcomes and ΔX the set of objective lotteries over X. If Receiver chooses f and state ω realizes, an outcome is generated by lottery f_ω. In the restaurant example above, for instance, there are two states and ordering a burger corresponds to the act $(Burger, Burger)$ (the same deterministic outcome in each state) and ordering a salad to $(Salad1, Salad2)$ (different outcomes in different states).

A *menu* is a finite set A of acts available to Receiver. The analysis involves different types of choice or preference primitives based on menus:

- *Menu preferences* $\succsim$ indicating Receiver's ex-ante ranking of menus; $A \succsim B$ means Receiver prefers committing to menu A over committing to menu B.

- *Random choices* ρ indicating Receiver's unconditional choice distributions from menus; $\rho^A(f) \in [0, 1]$ is the frequency with which Receiver chooses f from A.

- *State-contingent random choices* λ_ω indicating choice distributions in different states; $\lambda_\omega^A(f) \in [0, 1]$ is the frequency with which Receiver chooses f from A in state ω.

- *Choice correspondence* data c indicating which actions are chosen with positive probability; $c(A)$ is the support of ρ^A (or the union over all ω of the support of λ_ω^A).

Persuasion Representations of these primitives involve three parameters (μ, u, v): prior beliefs, Receiver's utility function and Sender's utility function, respectively. Such representations reflect Receiver's preferences or choices given that an (unobserved) Sender controls the information available to Receiver. At signal realizations, Receiver updates beliefs via Bayes' rule and chooses actions to maximize expected u; anticipating this, Sender chooses information to maximize expected v. In Persuasion Representations of $\succsim$, $A \succsim B$ indicates Receiver's expected utility from the interaction is greater at menu A than at menu B. In Persuasion Representations of ρ, distributions ρ_A match the distribution of Receiver's choices generated by Sender's optimal information structure at A; similarly, Persuasion Representations of λ require λ_ω^A to match the distribution generated by the interaction in state ω.

The paper develops several results for Persuasion Representations, organized as follows:

- *Identification (Section 3).* The parameters (μ, u, v) are identified by any of the above primitives, with one caveat: identification of μ from λ or c requires $u \not\approx v$, where $\approx$ indicates positive affine transformation.

- *Comparative Statics (Sections 4.1 & 4.2).* I show how Receiver's value of commitment and information varies with the degree of conflict with Sender. I characterize extreme cases such as $u \approx \pm v$ as well as smoother notions of "more-aligned" preferences.

- *Sophistication & Naivete (Section 4.3).* Comparing $\succsim$ to λ reveals whether Receiver correctly forecasts behavior or is optimistic/pessimistic regarding Sender's motives.

- *Axiomatic Foundations (Supplemental Appendix).* I provide an axiomatic characterization of Persuasion Representations for menu preferences $\succsim$ that, in turn, yields a characterization for c (hence ρ and λ). My approach takes μ as given in order to simplify the presentation but can be modified to drop this assumption.

The identification results show that Bayesian Persuasion interactions can be understood entirely from the perspective of Receiver: direct observation of Sender is not required to elicit parameter values—hence endogenous signals—and various types of Receiver data each suffice. The fact that c reveals both u and v is somewhat unusual relative to the literature, as such data do not involve preferences for commitment or even sophistication (correct forecasting of future behavior) on the part of the agent.

Given this identification result, most results in section 4 are established for $\succsim$ and c. For example, Proposition 1 shows that $\succsim$ satisfies *Preference for Flexibility*—$A \supseteq B$ implies $A \succsim B$—if and only if c satisfies *Sen's Condition* α—$A \subseteq B$ implies $c(B) \cap A \subseteq c(A)$—and that either property holds if and only if $u \approx v$ or $u \approx -v$. Thus, in Persuasion Representations, Sen's Condition α (Sen, 1971) for choice *from* menus is equivalent to Preference for Flexibility (Kreps, 1979) for choice *between* menus. My results reveal many such relationships between $\succsim$ and c. As a practical matter, the fact that Preference for Flexibility permits $u \approx -v$ is a significant departure from related models and points to nuanced relationships between temptation and commitment when selective attention is in play.

I also analyze Receiver's value of additional information. This involves a menu operation that simulates public signals: given A and a Blackwell experiment σ, the menu σA mixes acts in A so that it is *as if* Receiver (i) chooses from the original A, and (ii) before doing so, observes a signal from σ in addition to, and independently of, that generated by Sender.[7] Thus, σ acts as a lower bound on the experiments available to Sender to choose from. By construction, $\sigma A \supseteq A$, so such bounds provide flexibility to Receiver.[8] Proposition 1 shows that *Preference for Information*—$\sigma A \succsim A$ for all A and σ—is equivalent to Preference for Flexibility and that Sen's Condition α is equivalent to *Informational Sen's* α—$c(\sigma A) \cap A \subseteq c(A)$. Thus, my findings establish links between the value of commitment and of information: results involving preference for commitment have *equivalent* value-of-information analogues.[9]

[7]The menu σA is virtually identical to a construction of Blackwell (1951,1953), adapted to the domain of Anscombe-Aumann acts; Wang (2022) employs an equivalent such adaptation.

[8]As explained in section 4, however, acts in σA offer a kind of commitment power by giving Receiver the ability to delegate choices conditional on σ-signals to another party (eg, to the server at the restaurant).

[9]While relationships between dynamic (in)consistency and demand for information have been examined in

An important question in any dual-selves setting is whether the individual is sophisticated regarding their own future behavior. My approach, similar to that of Ahn et al. (2019), involves comparisons between ex-ante values of menus (preferences $\succsim$) and ex-post choices λ from menus. For sophisticated agents, the index v revealed by $\succsim$ matches that revealed by λ; otherwise, the agent may be optimistic or pessimistic regarding v. I provide full characterizations of these cases. Notably, λ (not c) turns out to be most useful for this exercise, as even sophisticated agents often violate $A \sim c(A)$; that is, Receiver's welfare is affected by unchosen alternatives (section 4.3 provides examples).

Finally, the axiomatic characterizations (see the Supplementary Appendix) establish how one might test whether Receiver choice data $\succsim$, ρ, λ or c are consistent with Bayesian Persuasion. The axioms also facilitate comparisons to other models. For example, preferences $\succsim$ in my model typically violate the *Desire for Commitment* axiom ($f \succsim A$ for some $f \in A$) of Dekel et al. (2009); instead, my model satisfies $A \succsim f$ for all $f \in A$. The Dekel et al. (2009) representation encapsulates several other models of temptation, making this distinction a key test for attention-based mechanisms of resolving intrapersonal conflict.

1.3 Related Literature

1. Bayesian Persuasion and Comparative Statics. My model is a decision-theoretic analogue of Bayesian Persuasion (Kamenica and Gentzkow, 2011). Crucially, I focus on Receiver's (not Sender's) choices and welfare and parameters (μ, u, v) are not exogenously specified but revealed by Receiver's behavior. My analysis involves comparisons between different persuasion games (menus or action sets), generating insights into how Receiver's welfare and behavior varies with the stakes, trade offs and degree of conflict with Sender.

Curello and Sinander (2022) establish rich and general results on the comparative statics of Bayesian Persuasion—specifically, they characterize conditions on model parameters under which Sender chooses a more informative (Blackwell, 1951) structure. This paper also studies comparative statics but the questions considered are different, as is the methodology. Changes to Receiver's welfare need not be due to Blackwell-comparable changes to information, so neither study nests the other.

2. Self-signaling and Strategic Ignorance. The behavioral interpretation developed in this paper incorporates elements of the self-signaling model of Bénabou and Tirole (2002). In their model, a time-inconsistent agent holds beliefs about the returns to costly effort and

the literature, direct connections between commitment and the value of information of the type considered here have not. Most decision-theoretic models of intrapersonal conflict do not involve menus of acts but rather of lotteries, and those that do have not utilized the σA construction.

may engage in selective information acquisition and/or costly memory retrieval to motivate such effort. Carrillo and Mariotti (2000) analyze a model of intrapersonal conflict where a present-biased individual sometimes chooses not to acquire freely-available information—strategic ignorance acts as a commitment device. Jakobsen (2021) examines this channel in arbitrary decision problems by studying Sender's preferences for information in Bayesian Persuasion.[10] Here, by providing justifications for tempting choices, information acquisition serves not as a commitment device but as an *enabler*.

3. Temptation, Demand for Commitment, Sophistication. Temptation and related behavior remains an active area of research.[11] Models typically generate demand for commitment via the combination of intrapersonal conflict and sophistication (awareness of the conflict); menu preferences directly express this demand and as such are the standard primitive used to characterize models. Here, menu preferences suffice but so do choice data ρ, λ, or c. The latter primitives do not express demand for commitment or require sophistication of the agent, providing new avenues to test for temptation problems and identify parameters.

4. Justification, Rationalization, Regret. In the behavioral interpretation, a choice is justified if a signal realization makes it u-maximal at posterior beliefs. Other models of justifiable choice, like Kalai et al. (2002), typically involve multiple rationales (preference orderings) and a choice is justified if it is maximal under at least one rationale.[12] The distribution of expected-utility preferences arising in my model (involving the same u but different posterior beliefs) is thus analogous to a set of rationales that varies endogenously with the menu.

Regret avoidance can also generate selective attention. Wang (2022) adapts the framework of Sarver (2008) to demonstrate individuals may prefer less information in order to avoid learning that past choices were ex-post suboptimal; thus, selective attention helps justify prior actions by suppressing new evidence against their desirability.

5. Random Choice and Inattention. Persuasion Representations involve a new mechanism generating random choice. Random utility models (Falmagne, 1978; Gul and Pesendorfer, 2006) cannot rationalize behavior generated by Persuasion Representations because the information chosen by Sender (hence, the distribution of expected utility functions governing

[10]The setup is similar to this paper but involves direct observation of Sender's informational preferences. Here, Sender's behavior is unobserved and must be inferred from Receiver's, so the model, results, and interpretation are quite different.

[11]In addition to an extensive theoretical literature, many recent studies approach the topic empirically or experimentally; see, for example, Toussaert (2018), Schilbach (2019), or Cobb-Clark et al. (2024). Table 1 of Carrera et al. (2022) summarizes over 30 recent studies on take-up of commitment contracts.

[12]See also Cherepanov et al. (2013), Lehrer and Teper (2011), and Ridout (2023).

Receiver's choices) varies with the menu of alternatives. For the same reason, the representations of Lu (2016) and Dekel and Lipman (2012) cannot rationalize choices generated by Persuasion Representations. Models of costly contemplation (Ergin and Sarver, 2010) or rational inattention (Ellis, 2018; Caplin and Dean, 2015) can rationalize Persuasion Representations only if one allows the cost of information to vary with the menu; if the cost function is menu-independent, increasing in the Blackwell order, and non-constant, the resulting model cannot rationalize Persuasion Representations.[13]

6. Identification and Characterization of Unobserved Signals. Several studies derive conditions under which choice data are consistent with rational choice under some information structure; see, for example, Dillenberger et al. (2014), Lu (2016), Azrieli and Rehbeck (2022), Rehbeck (2023), or Doval et al. (2024). These studies do not fully characterize sender-receiver communication models or analyze how the parameters of such models might be identified. In that sense, the closest work to this paper is Jakobsen (2021); as noted above, the contribution of that paper is quite different because it focuses on environments where Sender's informational preferences are directly observed.

2 Persuasion Representations

2.1 Framework

States, Outcomes, Acts

The outcome generated by Receiver's action depends on an exogenous state of the world. To capture this, I model actions as Anscombe-Aumann **acts** $f : \Omega \to \Delta X$ where:

- Ω is a finite set of $N \geq 2$ states (with generic members ω),
- X is a finite set of outcomes (generic members x, y), and
- ΔX is the set of lotteries over X (generic members p, q); lottery p delivers outcome x with probability $p(x)$.

Acts, or *state-contingent lotteries*, may be written as profiles $f = (f_\omega)_{\omega \in \Omega}$ where $f_\omega := f(\omega)$. In state ω, f returns lottery f_ω which in turn generates outcome x with probability $f_\omega(x)$. Constant acts $(p, \dots, p)$ are typically denoted p. Let F denote the set of all acts and $\mathcal{A}$ the

[13] Informally, one can construct menus where (i) perfect information is chosen by Sender in a Persuasion Representation, but (ii) the stakes are so low that the utility difference between perfect information and no-information does not outweigh the (menu-independent) cost of acquiring perfect information.

set of all finite, nonempty subsets of F. A set $A \in \mathcal{A}$ serves as an action set, or **menu**, of acts available for Receiver to choose from. Singleton menus $\{f\}$ are typically denoted f.

Lotteries and acts are equipped with standard mixing operations. In particular, $\alpha p + (1-\alpha)q$, where $\alpha \in [0,1]$, denotes a lottery r such that $r(x) = \alpha p(x) + (1-\alpha)q(x)$ for all x. This operation extends to acts by defining $\alpha f + (1-\alpha)g$ as the act h such that, for all $\omega \in \Omega$, $h_\omega = \alpha f_\omega + (1-\alpha)g_\omega$. These operations generalize to finite mixtures $\alpha_1 p^1 + \ldots + \alpha_n p^n$ or $\alpha_1 f^1 + \ldots + \alpha_n f^n$, where $\alpha_i \geq 0$ and $\alpha_1 + \ldots + \alpha_n = 1$, in the natural way.

Experiments and Signals

A Blackwell **experiment** is a matrix with $|\Omega| = N$ rows, finitely many columns, and entries in $[0,1]$ such that each row constitutes a probability distribution and no column consists entirely of zeros. Let $\mathcal{E}$ denote the set of all experiments, with generic members σ. Each column of an experiment represents a message that may be generated and each row a state-contingent distribution over messages. For example, the $N \times N$ identity matrix, denoted σ^*, associates a distinct message to each state and therefore represents **perfect information**.

An experiment can be expressed in terms of its columns. To do so, let S denote the set of all profiles $s = (s_\omega)_{\omega \in \Omega}$ of numbers $s_\omega \in [0,1]$ such that $s_\omega \neq 0$ for at least one $\omega \in \Omega$. Elements of S, **signals**, represent columns that may be present in an experiment. Abusing notation slightly, '$s \in \sigma$' indicates that s is a column of σ. A matrix $[s^1, \ldots, s^n]$ of signals is an experiment if and only if $s^1 + \ldots + s^n = e$, where $e = (1, \ldots, 1) \in S$ denotes an **uninformative signal** (or **uninformative experiment** since e qualifies as an experiment).

Signals and experiments yield additional mixture operations on acts. If $s \in S$, let $sf + (1-s)g$ denote the act h such that $h_\omega = s_\omega f_\omega + (1-s_\omega)g_\omega$; this operation is similar to the α-mixture of f and g defined above but allows potentially different weights s_ω to be applied in different states ω. More generally, if $\sigma = [s^1, \ldots, s^n]$ is an experiment, $s^1 f^1 + \ldots + s^n f^n$ denotes the act h such that $h_\omega = s^1_\omega f^1_\omega + \ldots + s^n_\omega f^n_\omega$; h_ω is a well-defined lottery because σ is a Blackwell experiment and, thus, $s^1_\omega + \ldots + s^n_\omega = 1$.

Choice Primitives

My analysis involves different kinds of preference or choice data based on menus:

1. **Menu preferences** $\succsim$ over $\mathcal{A}$ where $A \succsim B$ means Receiver prefers committing to menu A over committing to menu B.

2. **Random choice** data $\rho = (\rho^A)_{A \in \mathcal{A}}$ where ρ^A is a probability distribution over A and $\rho^A(f) \in [0,1]$ is the probability Receiver chooses f from A.

3. **State-contingent random choice** data $\lambda = (\lambda_\omega^A)_{\omega \in \Omega, A \in \mathcal{A}}$ where λ_ω^A is a probability distribution over A and $\lambda_\omega^A(f) \in [0,1]$ is the probability Receiver chooses f from A in state ω.[14]

4. **Choice correspondence** data $c : \mathcal{A} \to \mathcal{A}$ where $c(A) \subseteq A$ is the set of acts Receiver chooses from A with positive probability (in at least one state).

Note that c can be derived from ρ or λ but that no other pairs of primitives are nested in such a way. Most of my analysis treats the primitives separately (eg, does not require the analyst to observe *both* $\succsim$ and ρ). The only exception is section 4.3, which combines $\succsim$ with λ to study the agent's sophistication.

2.2 Persuasion Representations

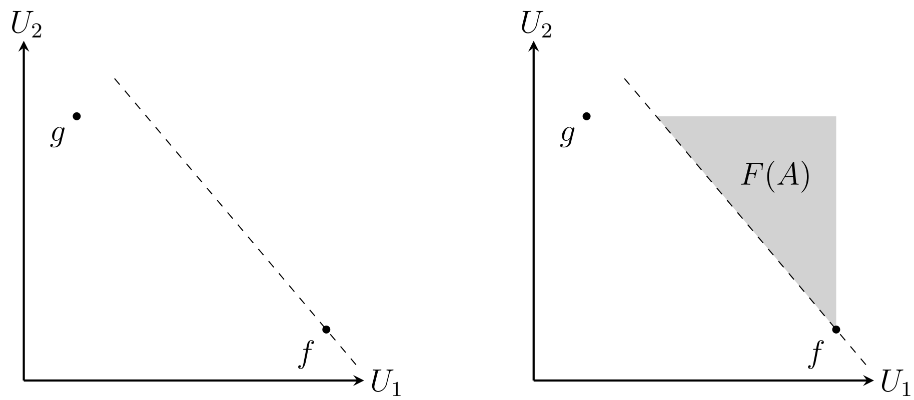
This section defines Persuasion Representations for each type of primitive described above. Let $\Delta\Omega$ denote the set of probability distributions over Ω. The representations involve three parameters, (μ, u, v), where $\mu \in \Delta\Omega$ is a prior and $u, v : X \to \mathbb{R}$ are utility indices for Receiver and an (unobserved) Sender, respectively. The notation $u \approx u'$ indicates positive affine transformation: there exist $\alpha > 0$, $\beta \in \mathbb{R}$ such that $u'(x) = \alpha u(x) + \beta$ for all $x \in X$.

For any $\hat{\mu} \in \Delta\Omega$ and state ω, let $\hat{\mu}_\omega$ denote the probability of state ω. Throughout, prior beliefs μ have full support and u, v are non-constant. The indices u, v apply to lotteries as follows: if $p \in \Delta X$, then $u(p) := \sum_{x \in X} u(x)p(x)$ is Receiver's expected utility of p; similarly, $v(p)$ denotes Sender's expected utility. For acts, Receiver's expected utility is given by $U : F \to \mathbb{R}$ where $U(f) := \sum_{\omega \in \Omega} u(f_\omega)\mu_\omega$. More generally, for any signal $s \in S$, let $U^s(f) := \sum_{\omega \in \Omega} u(f_\omega)s_\omega\mu_\omega$; this represents Receiver's expected utility conditional on s.[15] Note that $U^e = U$. Replacing u with v leads to functions $V, V^s : F \to \mathbb{R}$ representing Sender's expected utility conditional on signal realizations.

In a persuasion game at A, Receiver chooses an act from A after observing a signal generated by Sender. Sender provides information by freely selecting an experiment σ and committing to revealing its signal realization. Sender correctly forecasts Receiver's signal-contingent choices and constructs σ to maximize his own expected payoff. The agents share a common prior but typically differ in their preferences over outcomes generated by the chosen act(s); this is the source of conflict and it varies with the menu—some menus consist of acts for which Sender and Receiver largely agree on the best course of action, others less so.

[14] Note that by conditioning on ω, choice data λ_ω^A indicates that the analyst (not Receiver) knows ω.

[15] This holds because the Bayesian posterior of μ at s assigns probability $\frac{s_\omega \mu_\omega}{s \cdot \mu}$ to state ω, where $s \cdot \mu = \sum_{\omega' \in \Omega} s_{\omega'}\mu_{\omega'}$. Thus, expected utility conditional on s is $\sum_{\omega \in \Omega} \frac{u(f_\omega)s_\omega\mu_\omega}{s \cdot \mu}$. The function U^s multiplies this value by the constant $s \cdot \mu > 0$ and therefore represents the same ranking of acts.



(a) Menu $A = \{f, g\}$ in utility space (b) $F(A)$ for $A = \{f, g\}$

Figure 1: Construction of $F(A)$ for two states and $A = \{f, g\}$. The dashed line is Receiver's indifference curve through f; its slope is determined by μ and indicates that f is prior-optimal. Therefore, f is an induced act: it corresponds to Sender choosing $\sigma = e$ (no information). The upper-right corner of $F(A)$ is the induced act (f_1, g_2) corresponding to perfect information. The upper-left point is induced by maximizing the probability of posterior beliefs that make Receiver indifferent between f and g. In general, $F(A)$ is a polytope in F bounded by (and passing through) Receiver's prior-optimal act.

To begin, it is useful to express the range of state-contingent lotteries that can be achieved by varying the information available to Receiver. At menu A, an experiment σ transforms into an act as follows. Fix a state ω. In this state, σ generates a distribution over signals ($s \in \sigma$ is generated with probability s_ω) and at every $s \in \sigma$ Receiver chooses a U^s-optimal act $f^s \in A$. In state ω, this act delivers a lottery f^s_ω. Thus, the state-contingent distribution over signals becomes a distribution over lotteries, which reduces to a single lottery in the natural way. Repeating this procedure for each state yields an **induced act**: a state-contingent lottery over outcomes generated by Receiver's choices under information σ at menu A.

The above procedure associates a unique induced act to an experiment σ if there is a unique U^s-optimal act $f^s \in A$ for each $s \in \sigma$. If there are multiple U^s-optimal acts for some s, different tie-breaking selections produce different induced acts. To capture the full range of possibilities, the set of induced acts at A is defined as

$$F(A) := \left\{ \sum_{s \in \sigma} s f^s : \sigma \in \mathcal{E},\ f^s \in \text{co}(A),\ U^s(f^s) \geq U^s(g)\ \forall g \in A \right\}, \tag{1}$$

where $\text{co}(A)$ is the convex hull of A. This set contains all induced acts generated by varying both σ and Receiver's tie-breaking behavior: if Receiver finds two or more acts optimal at s, then f^s is permitted to be any convex combination of those acts. Note that only parameters (μ, u) are needed to construct $F(A)$. Figure 1 provides an illustration.

Since Sender anticipates Receiver's choices, Sender's choice of information at A is effectively a choice from $F(A)$. This leads to the following definition for menu preferences:

Definition 1. Parameters (μ, u, v) constitute a **Persuasion Representation** for $\succsim$ if u, v are non-constant, μ has full support, and the function $U : \mathcal{A} \to \mathbb{R}$ given by

$$U(A) := \max U(f) \;\; \text{subject to} \;\; f \in \operatorname*{argmax}_{g \in F(A)} V(g) \tag{2}$$

represents $\succsim$, where $U(f) = \sum_\omega u(f_\omega)\mu_\omega$ and $V(f) = \sum_\omega v(f_\omega)\mu_\omega$.[16]

In a Persuasion Representation for $\succsim$, Sender correctly forecasts Receiver's signal-contingent choices from A and selects an information structure, hence an induced act $f \in F(A)$, that maximizes his own expected utility. When evaluating A, Receiver correctly forecasts Sender's choice and assigns value $U(f)$ to A, where f is the induced act associated with the chosen information structure. Thus, $U(A)$ is Receiver's ex ante expected utility from the interaction at A. This is well-defined because $F(A)$ is compact (see the Supplementary Appendix).

Implicitly, formula (2) makes two assumptions about tie-breaking behavior. First, the requirement that $f \in \operatorname{argmax}_{g \in F(A)} V(g)$ means that if multiple acts maximize U^s at some s, Sender expects Receiver to select a V^s-maximal act among the U^s-maximizers. This is the standard "Sender-preferred" tie-breaking rule in the Bayesian Persuasion literature and it ensures existence of a Sender-optimal information structure at every A. On a technical level, it emerges from maximization over $F(A)$ because that set includes all possible induced acts that come about by varying both information and Receiver's tie-breaking selections.

Second, the "max" in (2) means that if Sender finds multiple information structures optimal at A, he selects from such structures a Receiver-optimal one. Formally, let

$$V^\sigma(A) := \max \sum_{s \in \sigma} V^s(f^s) \text{ subject to } f^s \in \operatorname*{argmax}_{f \in A} U^s(f)$$

and

$$U^\sigma(A) := \sum_{s \in \sigma} U^s(f^s) \text{ where } f^s \in \operatorname*{argmax}_{f \in A} U^s(f).$$

These functions capture Sender's and Receiver's value, respectively, of information σ at A: if Sender chooses σ, he expects payoff $V^\sigma(A)$ and Receiver expects $U^\sigma(A)$. Clearly, $V^\sigma(A)$ in-

[16] See the Supplementary Appendix for an axiomatic characterization. Note that it is straightforward to ensure u is non-constant (Axiom 1 does so) but that if v is constant, Receiver's behavior is indistinguishable from the $u \approx v$ case. Thus, for both for menu preferences and the other primitives, it is without loss to assume v is non-constant as well.

corporates the Sender-preferred tie-breaking rule described above. The "Receiver-preferred" rule implied by (2) means that if multiple experiments σ maximize $V^\sigma(A)$, Sender breaks the tie by maximizing $U^\sigma(A)$. The following theorem makes this explicit by re-expressing Persuasion Representations for $\succsim$ in a more familiar form; I omit the straightforward proof.

Theorem 1. *The function U defined by (2) above satisfies*

$$U(A) = \max_{\sigma} U^\sigma(A) \text{ subject to } \sigma \in \operatorname*{argmax}_{\sigma' \in \mathcal{E}} V^{\sigma'}(A). \tag{3}$$

Theorem 1 states that formulas (2) and (3) coincide; (3) more-directly expresses Receiver's payoff as the result of Sender's information design problem but, as we shall see, (2) facilitates comparisons to related models and is useful for deriving various results, including the axiomatization (see the Supplementary Appendix).

Persuasion Representations for ρ, λ, and c do not involve Receiver's ex ante value $U(A)$ but rather the actual choices from A stemming from Sender's information structure. Let $\mathcal{E}^*(A) \subseteq \mathcal{E}$ denote the set of solutions to the maximization problem (3). An experiment $\sigma \in \mathcal{E}^*(A)$ is A**-minimal** if there is no $\sigma' \in \mathcal{E}^*(A)$ such that σ' is a garbling of σ and $\sigma' \neq \sigma$. Given A, a **behavioral strategy** is a profile $\beta^A = (\beta^{A,s})_{s \in S}$ such that $\beta^{A,s} \in \Delta A$ for all $s \in S$; that is, $\beta^{A,s}$ is a distribution of choices from A at signal s.

Definition 2. Parameters (μ, u, v), where u, v are non-constant and μ has full support, constitute a **Persuasian Representation** for ρ if for every A there is a behavioral strategy β^A and A-minimal experiment $\sigma \in \mathcal{E}^*(A)$ such that

(i) for all $s \in S$,

$$\operatorname{supp}(\beta^{A,s}) = \operatorname*{argmax}_{f} V^s(f) \text{ subject to } f \in \operatorname*{argmax}_{g \in A} U^s(g), \text{ and}$$

(ii) for all $f \in A$, $\rho^A(f) = \sum_{s \in \sigma} (s \cdot \mu) \beta^{A,s}(f)$.

Informally, parameters (μ, u, v) constitute a Persuasion Representation for ρ if, for every A, choices frequencies ρ^A coincide with those generated by Sender's chosen experiment and Receiver's signal-contingent choices for that experiment; these choices must be optimal given parameters (μ, u, v). Part (i) of Definition 2 requires that, for every $s \in \sigma$, Receiver chooses the act(s) that are consistent with his own optimization and the Sender-preferred tie-breaking criterion. Part (ii) requires that the observed probability of choosing f from A, $\rho^A(f)$, coincides with the total probability of choosing f given σ and β^A; in particular, $s \cdot \mu$ (the dot product) is the total probability of generating signal $s \in \sigma$ under prior μ, and $\beta^{A,s}(f)$ is the probability of choosing f conditional on signal s.

Since the chosen experiment σ is a member of $\mathcal{E}^*(A)$, it satisfies the Receiver-preferred tie-breaking criterion described above. It is also required to be A-minimal. While neither agent's payoffs or incentives are affected by the A-minimality requirement, it implies that if e (no information) is both Sender- and Receiver-optimal, then e is chosen by Sender. This simplifies the statements and proofs of several results.

Definition 3. Parameters (μ, u, v), where u, v are non-constant and μ has full support, constitute a **Persuasian Representation** for λ if for every A there is a behavioral strategy β^A and A-minimal experiment $\sigma \in \mathcal{E}^*(A)$ such that

(i) for all $s \in S$,

$$\text{supp}(\beta^{A,s}) = \underset{f}{\text{argmax}}\, V^s(f) \text{ subject to } f \in \underset{g \in A}{\text{argmax}}\, U^s(g), \text{ and}$$

(ii) for all $f \in A$ and $\omega \in \Omega$, $\lambda^A_\omega(f) = \sum_{s\in\sigma} s_\omega \beta^{A,s}(f)$.

The definition of a Persuasion Representation for λ is nearly identical to that of ρ. The only difference is that *state-contingent*, as opposed to *total*, choice frequencies must agree with those generated by the persuasion game with parameters (μ, u, v); condition (ii) reflects this.

Definition 4. Parameters (μ, u, v), where u, v are non-constant and μ has full support, constitute a **Persuasian Representation** for c if for every A there is a behavioral strategy β^A and A-minimal experiment $\sigma \in \mathcal{E}^*(A)$ such that

(i) for all $s \in S$,

$$\text{supp}(\beta^{A,s}) = \underset{f}{\text{argmax}}\, V^s(f) \text{ subject to } f \in \underset{g \in A}{\text{argmax}}\, U^s(g), \text{ and}$$

(ii) $c(A) = \bigcup_{s\in\sigma} \text{supp}(\beta^{A,s})$.

Intuitively, parameters (μ, u, v) constitute a Persuasion Representation for c if, for every A, $c(A)$ coincides with the support of ρ^A where ρ has a Persuasion Representation with parameters (μ, u, v). Alternatively, $c(A)$ coincides with $\bigcup_{\omega\in\Omega} \text{supp}(\lambda^A_\omega)$ where λ has a Persuasion Representation. Either way, $c(A)$ contains all acts in A that are chosen with positive probability in at least one state; any additional frequency information is discarded.

Before concluding this section with some comments on Persuasion Representations, a brief discussion of condition (i) in Definitions 2–4 is in order. Condition (i) requires that, at each signal realization, every act in A that satisfies the Sender-preferred tie-breaking rule is chosen with positive probability. There are other ways of refining choices, but the advantage of (i), together with the A-minimality requirement, is that special cases of the model reduce to familiar representations in decision theory. For example, if $u \approx -v$, Sender chooses no-information and Receiver simply chooses based on his prior. By (i), then, $c(A)$ consists of *all* prior-optimal acts in A, as in standard choice models, rather than some arbitrary subset of optimal acts. For the purposes of this paper, such well-behaved special cases are not strictly necessary but allow a cleaner exposition and simpler proofs; this is the main reason for imposing tie-breaking conventions beyond the standard Sender-preferred criterion.

2.2.1 Persuasion Representations: Comments

The Behavioral Interpretation. As explained in the introduction, the behavioral interpretation involves two competing forces within the individual: consumption choices must maximize u conditional on beliefs but the agent allows temptation (utility v) to influence attentiveness to freely-available information. Choices at signal realizations s maximize U^s and are therefore "justified" even though the experiment generating those signals is designed to make choices increase v-utility. In essence, the agent engages in *motivated attention* in order to obtain justifications for choosing tempting options; the need for such justifications is a form of self-control.

When is information obtained? For menu preferences $\succsim$, the representation values menus according to their expected u-utility from the interaction; v enters the picture only when choosing *from* a menu. Therefore, information acquisition occurs after selecting a menu but before, or during, the time of consumption (eg, the agent looks at the menu, asks the server a question, then chooses an item based on the response). This is consistent with temptation being a transient phenomenon: the agent enters a "hot" state when confronted with tempting options but is otherwise in a "cold" state where only u matters. Note, however, that Persuasion Representations of ρ, λ or c need not involve such transience because they make no claims about the agent's ex-ante value of menus. An agent who anticipates having to make choices from A may acquire information well in advance and could even be in a persistent "hot" state where v guides attention (eg, choosing news sources to pay attention to in the months leading up to an election).

Strotz vs Persuasion. The representation of Definition 1 is similar to a Strotzian representation (Strotz, 1955). In the notation of this paper, such representations take the form

$$W_{Strotz}(A) := \max U(f) \text{ subject to } f \in \operatorname*{argmax}_{g \in A} V(g). \tag{4}$$

A standard interpretation is that one self, with utility U, anticipates their future-self choosing from A via maximization of V. In other words, the agent's utility function over acts *changes* from U to V. The Persuasion Representation of Definition 1 has a similar mathematical structure, with one key difference: the future self chooses an act not from A, but from $F(A)$. This captures the idea that temptation influences future choice through information acquisition: signal-contingent choices must maximize U, so selective attention can only generate acts in $F(A)$. Like the Gul and Pesendorfer (2001) model, described below, mine is not a model where U changes to V; instead, the agent experiences temptation (a desire to increase V-utility) but exercises a form of self-control (only choosing U^s-maximal options at signal realizations s) in response.

The Cost of Temptation. In the Gul and Pesendorfer (2001) model, the decision maker may resist temptation but suffer a cost of self-control from doing so. In the notation of this paper, their representation may be written as

$$W_{GP}(A) = \max_{f \in A} U(f) - \left[\max_{g \in A} V(g) - V(f) \right].$$

Thus, when choosing among acts $f \in A$, the agent weighs U-utility against $\max_{g \in A} V(g) - V(f)$, the cost of self-control. Depending on A, the agent may choose a U-maximal option, a V-maximal option or, more typically, a "compromise" option.

In my model, the agent behaves stochastically: at a typical menu A, Receiver sometimes chooses tempting options and sometimes not—the distribution of choices is, in effect, the "compromise" option. There is no cost of self-control, but Receiver effectively suffers a cost of *temptation* given by $U^{\sigma^*}(A) - U(A)$ (equivalently, $U^{\sigma^*}(A) - U^{\sigma}(A)$, where $\sigma \in \mathcal{E}^*(A)$). Intuitively, perfect information σ^* is freely available but temptation makes the agent acquire some $\sigma \in \mathcal{E}^*(A)$ instead, resulting in lower expected U-utility.

The Role of Sophistication. Most decision-theoretic models of intrapersonal conflict are characterized in terms of ex-ante menu preferences $\succsim$. Crucially, such representations involve forecasting of future behavior. For example, Strotzian representations of the form (4) require the ex-ante agent, with utility U, to anticipate future choices made with V; in this sense, the function V in the representation reflects the initial self's *beliefs* about future tastes.

Similarly, Persuasion Representations of $\succsim$ require Receiver to forecast Sender's choice of information; thus, the index v reflects Receiver's beliefs about Sender's tastes.

In contrast, Persuasion Representations of ρ, λ, and c do not involve such forecasting or beliefs—they merely express Receiver's choices resulting from the interaction. Consequently, the index v in such representations reflects the true preferences of Sender. Comparisons between $\succsim$ and (say) λ, therefore, can detect whether the ex-ante agent is sophisticated (holds correct beliefs about v) or naive (optimistic or pessimistic about v); see section 4.3.

Cheap Talk vs Persuasion. Other communication models, like cheap talk (Crawford and Sobel, 1982), are potentially amenable to decision-theoretic analysis and the behavioral interpretation studied in this paper. Cheap talk does not involve the oft-criticized assumption of Sender commitment power that lies at the core of Bayesian Persuasion models, but typically yields multiple qualitatively-distinct equilibria. The commitment assumption of Bayesian Persuasion mostly eliminates this problem, leaving only particular tie-breaking issues that are fairly commonplace in decision-theoretic models. The commitment assumption is also more appealing in the behavioral interpretation, where signals are the result of selective attention: the agent chooses an information source and directly observes its output.

3 Identification

This section establishes identification results for each type of choice primitive; section 3.1 provides a proof sketch and illustrates methods to elicit parameter values from the primitives.

Theorem 2. *If (μ, u, v) and (μ', u', v') are Persuasion Representations of a preference $\succsim$ on $\mathcal{A}$, then $u \approx u'$, $v \approx v'$, and $\mu = \mu'$.*

Theorem 2 states that Persuasion Representations of menu preferences are unique: if $\succsim$ has a Persuasion Representation, there is a unique prior μ and unique (up to positive affine transformation) utility indices u, v for which (μ, u, v) constitute a Persuasion Representation of $\succsim$. Thus, menu preferences $\succsim$, alone, are sufficient to identify all parameters. An analogous result holds for representations of random choice rules:

Theorem 3. *If (μ, u, v) and (μ', u', v') are Persuasion Representations of a random choice rule ρ, then $u \approx u'$, $v \approx v'$, and $\mu = \mu'$.*

A slightly weaker result holds for λ and c:

Theorem 4. *If* (μ, u, v) *and* (μ', u', v') *are Persuasion Representations of either a state-contingent random choice rule* λ *or a choice correspondence* c*, then: (i)* $u \approx u'$ *and* $v \approx v'$*, and (ii) if* $u \not\approx v$*, then* $\mu = \mu'$*.*

Theorem 4 states that u and v are identified from λ or c but that μ is identified by either primitive only if $u \not\approx v$; that is, if there is some conflict between Sender and Receiver. Since c is nested by ρ and λ, this means that outside the hairline case $u \approx v$, both ρ and λ contain more information than is actually required to uniquely identify all three parameters. For this reason, most subsequent results in the paper are established only for $\succsim$ and c.

To see what goes wrong when $u \approx v$, observe that such preferences make Sender choose perfect information at every menu A.[17] Consequently, in state ω, Receiver chooses precisely those acts $f \in A$ for which $u(f_\omega) \geq u(g_\omega)$ for all $g \in A$. These choices depend on ω but not the probability μ_ω with which ω realizes. Thus, $c(A)$ does not depend on μ. Similarly, when $u \approx v$, state-contingent choices λ_ω do not reveal anything about μ unless one makes arbitrary assumptions about how tie-breaking varies with μ.[18]

The main takeaway of Theorems 2–4 is that full identification can be achieved using any of the four types of preference or choice data, subject to the restriction $u \not\approx v$ for λ and c. Since (μ, u, v) are identified, so is Sender's choice of information—at any A, the optimal information structure(s) can be derived from (μ, u, v). Applying the behavioral interpretation, this means "true" utility u, temptation utility v and the information acquired to rationalize choices are revealed by standard choice data; to my knowledge, this is the first general result on identification in intrapersonal signaling models. Interestingly, ρ and λ provide more information than is actually required: simply knowing which acts are chosen with positive probability, as recorded by c, is enough to deduce (μ, u, v), making measurement error potentially less problematic. More broadly, the fact that consumption data like ρ, λ or c can reveal and identify temptation-style behavior is somewhat unusual; commitment preferences $\succsim$ tend to be the standard approach and are often required for such analysis.[19]

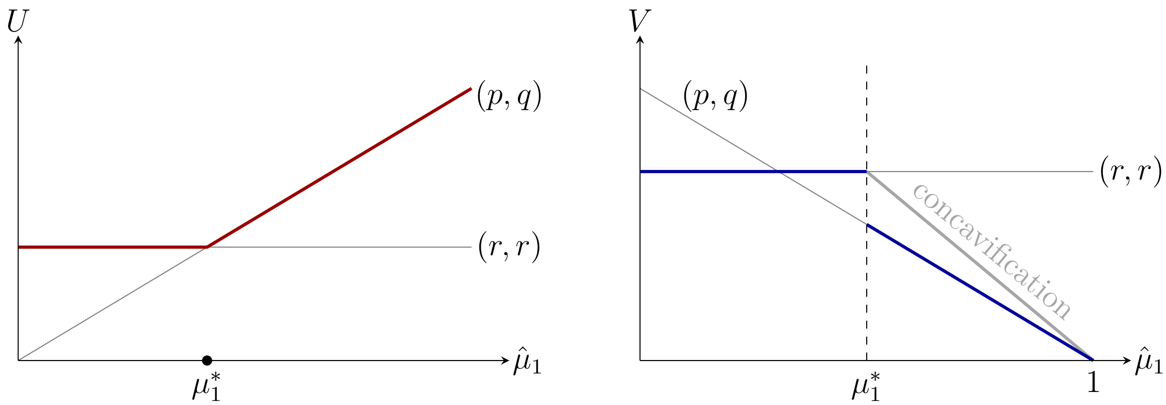
[17]More precisely, perfect information is Sender-optimal at every A; Sender may choose coarser information in some menus due to the A-minimality requirement, but this does not affect the argument.

[18]For example, one could specify a menu A where, in state ω_1, there is a tie between two acts and Receiver chooses one of them with probability μ_{ω_1}. If Receiver employs such a tie-breaking criterion and the analyst knows which particular act f is chosen with probability μ_{ω_1}, then $\lambda^A_{\omega_1}(f)$ reveals μ_{ω_1}. However, there is no reason to suspect that Receiver would correlate tie-breaking selections with μ or, if he did, that the analyst would know the correlation structure.

[19]Indeed, elicitation of commitment preferences is the standard approach in field and lab settings (see section 1.3 for references). In their survey of the temptation literature, Lipman and Pesendorfer (2013) argue that "choice over menus is useful if we want to study phenomena that the consumption choice alone cannot reveal. Specifically, it is useful to study choice affected by temptations."

(a) Eliciting u (b) Eliciting v

Figure 2: Identifying u and v.



(a) Receiver's value (b) Sender's value

Figure 3: Identifying μ from c.

3.1 Eliciting Parameters

Given lotteries p and q, a menu A is a pq**-bet** if there is a nonempty $E \subsetneq \Omega$ such that $A = \{pEq, qEp\}$, where $h = p'Eq'$ satisfies $h_\omega = p'$ for $\omega \in E$ and $h_\omega = q'$ for $\omega \notin E$.

Menu Preferences

Receiver's parameters (μ, u) are easily identified by considering the restriction of $\succsim$ to singleton menus. In particular, $\{f\} \succsim \{g\}$ if and only if $U(f) \geq U(g)$, so the Anscombe-Aumann theorem applies and μ and u are identified.

To elicit v, the key is to determine which pairs (p, q) of lotteries are ranked the same way by u and v. Figure 2b illustrates the idea. For any p, the index u gives a linear indifference curve through p. To identify v, it is enough to determine the slope of v's indifference curve through p and the direction of increasing utility. If $A_{pq} = \{pEq, qEp\}$ is a pq-bet where $u(p) > u(q)$, Sender either agrees with the ranking ($v(p) \geq v(q)$) or disagrees ($v(p) < v(q)$). If Sender agrees, he chooses perfect information and so $A_{pq} \sim p$; otherwise, Sender disagrees and chooses no-information, yielding $p \succ A_{pq}$. Thus, fixing p and eliciting all q such that $A_{pq} \sim p$ reveals the indifference curve for v through p and the direction of increasing utility.

Choice Distributions & Correspondences

For ρ, λ and c, elicitation is slightly more involved. The first step is to identify u by analyzing choices from menus $\{p, q\}$ of constant acts. Fixing p, we have $c(\{p, q\}) = p$ if $u(p) > u(q)$. If $u(p) = u(q)$, the representation breaks the tie in favor of v. Thus, $p \in c(\{p, q\})$ if and only if $u(p) > u(q)$ or both $u(p) = u(q)$ and $v(p) \geq v(q)$. As illustrated in Figure 2a, fixing p and eliciting all q such that $p \in c(\{p, q\})$ reveals the lower contour set of p for u; in particular, the closure of the set of all such q is the weak lower contour of u through p. Given u, one can identify v using similar techniques developed for menu preferences: if $u(p) > u(q)$, then (outside of hairline cases) $c(A_{pq}) = A_{pq}$ if and only if $v(p) \geq v(q)$.[20]

To see how μ may be identified from c, consider the two-state case and suppose $u \not\approx v$ and $u \not\approx -v$ (the proof provided in the appendix is different and only requires $u \not\approx v$). With such preferences, there are lotteries p, q, r such that the menu $A = \{(p, q), (r, r)\}$ gives rise to the value functions depicted by Figure 3; the horizontal axis' contain all possible posterior beliefs (ordered by the weight assigned to state 1) and the values are those induced by Receiver's choices from A at those beliefs.[21] At μ^*, Receiver is indifferent between the two acts, creating a discontinuity in Sender's value function. If prior beliefs satisfy $\mu_1 \leq \mu_1^*$, Sender chooses no-information; consequently, $c(A) = \{(r, r)\}$. If instead $\mu_1 > \mu_1^*$, Sender maximizes his payoff by choosing an experiment yielding two posteriors: one at μ^*, the other at $\hat{\mu}_1 = 1$. Consequently, $c(A) = A$ since Receiver chooses (r, r) at μ^* and (p, q) at $\hat{\mu}_1 = 1$. Thus, $c(A)$ indicates whether $\mu_1 \leq \mu_1^*$ or $\mu_1 > \mu_1^*$, providing objective information about μ since μ^* is pinned down by u. It follows that μ can be identified by examining $c(A)$ for all binary menus A—for example, by moving r along its v-indifference curve and thereby perturbing only $u(r)$ and, thus, μ_1^*.

4 Comparative Statics from Receiver's Perspective

This section analyzes comparative statics regarding the degree of conflict between Sender and Receiver. In contrast to most of the persuasion literature, the focus is on how the interaction affects Receiver's (not Sender's) choices and welfare.

The characterizations take two forms: (i) how Receiver's ex ante value and subsequent choices vary with increased flexibility, and (ii) how such values and choices vary with additional public information. The results thereby establish tight links between Receiver's value

[20] To circumvent tie-breaking issues for the case $\mu(E) = 1/2$, where $A_{pq} = \{pEq, qEp\}$, the proof uses a more general class of menus than pq-bets; see the appendix for details.

[21] The desired lotteries exist because when $u \not\approx v$ and $u \not\approx -v$, one may fix indifference curves (utility levels) for one agent and move along them to set utility levels for the other; see Lemma 1 in the appendix.

of flexibility and of information in persuasion models. While analysis of (i) involves comparisons between $\subseteq$-comparable menus, analysis of (ii) employs an operator on menus that simulates public information. For any A and σ, let

$$\sigma A := \left\{ \sum_{s\in\sigma} s f^s : f^s \in A \right\}.$$

The menu σA simulates an environment where Receiver chooses from A but is able to condition this choice on the realization from σ in addition to the signal generated by Sender.[22] Sender recognizes this but cannot correlate realizations from his chosen experiment with those of σ. The example below provides concrete illustrations.

Example 2. Suppose there are two states and let $A = \{f, g\}$. If $\sigma = \sigma^*$ (the identity matrix), then σ perfectly reveals the state and $\sigma A = \{(f_1, f_2), (g_1, f_2), (f_1, g_2), (g_1, g_2)\}$. Suppose $u(f_1) > u(g_1)$ and $u(g_2) > u(f_2)$; this means neither act dominates the other and, in particular, that f is preferred in state 1 and g is preferred in state 2. At σA, then, Receiver chooses $f^* := (f_1, g_2)$ regardless of the information Sender provides. Consequently, $U(\sigma A) = U(f^*) = u(f_1)\mu_1 + u(g_2)\mu_2$, the value of A under perfect information.

If instead σ is a noisy structure $\sigma = [s, t]$, we have $\sigma A = \{sf+tf, sf+tg, sg+tf, sg+tg\} = \{f, sf + tg, sg + tf, g\}$. If Receiver prefers f at s and g at t, his prior-optimal act in σA is $sf + tg$; this represents the average state-contingent lottery for Receiver if Sender provides no additional information, thereby forming a lower bound for Receiver's welfare in the persuasion game with public information: $U(\sigma A) \geq U(sf + tg)$.[23] ♦

When Sender and Receiver are different people, σA might represent an environment where Receiver chooses from A but has access to additional public information σ; alternatively, σ may represent a lower bound on Sender's choice of information at A, making the interaction a constrained version of Bayesian Persuasion.[24]

For intrapersonal conflict, other interpretations of σA may be more appropriate. For example, σ might represent information that is too salient to ignore; features of the choice environment like the framing of information can affect such salience. More generally, σA

[22] Note that σA is virtually identical to the set of risk vectors central to the Blackwell (1951,1953) information order. In the modern formulation due to de Oliveira (2018), the set of risk vectors for (A, σ) is formed by considering all compositions of σ (viewed as a map $\Omega \to \Delta S$) with strategies $\beta : S \to \Delta A$. My treatment differs by taking A to be a set of Anscombe-Aumann acts (not abstract actions) and by restricting to pure strategies $\beta : S \to A$, resulting in a finite menu σA. Wang (2022) employs an equivalent construction.

[23] Note that this bound is attained only if Sender chooses no-information. Introducing public information affects Sender's incentives, so it need not be the case that $U(\sigma A) \geq U(A)$ (see Proposition 1 below).

[24] For example, such bounds could be due to laws requiring minimal degrees of transparency in advertising.

represents a situation where Receiver can commit to conditioning choices on realizations from σ in addition to whatever signal their attention provides. In addition to psychological mechanisms like willpower, such commitment might be achieved through delegation. In the restaurant example of the introduction, for instance, Receiver might ask the server to acquire a signal about the salad and to enter an order conditional on the signal; acts of the form $sf + tg$ in Example 2 above provide just that kind of delegation. Note that Sender chooses information knowing such delegation opportunities are available—the menu σA is known to both selves. Thus, σA could represent a situation where the server presents Receiver with the menu A *and* an offer to delegate the choice; Sender's choice of information takes that offer into account. Despite the various interpretations, I will simply refer to menus σA as settings with additional public information.

4.1 The Value of Flexibility and Information

This section characterizes special cases of the model involving standard rationality postulates. Under what circumstances does Receiver benefit from increased flexibility or from public information, and how is this reflected in choice data?

Proposition 1. *Suppose (μ, u, v) represents $\succsim$ and c. Then $u \approx v$ or $u \approx -v$ if and only if any of the following conditions hold:*

(i) $\succsim$ satisfies Preference for Flexibility: $A \supseteq B$ implies $A \succsim B$.

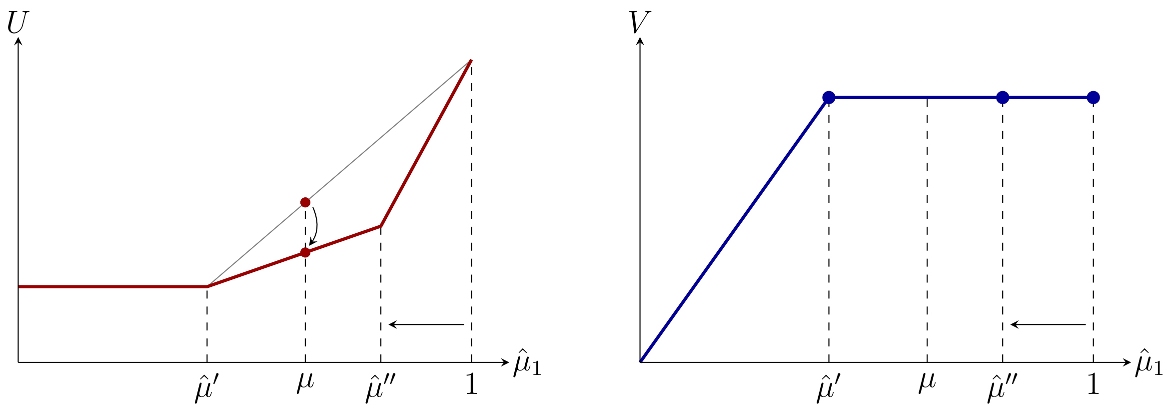
(ii) c satisfies Sen's condition α: $A \subseteq B$ implies $c(B) \cap A \subseteq c(A)$.

(iii) $\succsim$ satisfies Preference for Information: for all σ and A, $\sigma A \succsim A$.

(iv) c satisfies Informational Sen's α: $c(\sigma A) \cap A \subseteq c(A)$.

Proposition 1 states that in Persuasion Representations, Receiver's preferences and choices satisfy standard rationality postulates if and only if one of two extreme cases holds: the conflict between Sender and Receiver is either non-existent ($u \approx v$) or total ($u \approx -v$). To aid the discussion, note that if $u \approx v$, perfect information is Sender-optimal at all menus, reducing the representation of $\succsim$ to $\overline{U}(A) := \sum_{\omega\in\Omega} \max_{f\in A} u(f_\omega)\mu_\omega$ and that of c to $\overline{c}(A) = \bigcup_{\omega\in\Omega} \text{argmax}_{f\in A}\, u(f_\omega)$. If $u \approx -v$, Sender chooses e (no information) at all menus, reducing the representations to $\underline{U}(A) := \max_{f\in A} U(f)$ and $\underline{c}(A) = \text{argmax}_{f\in A}\, U(f)$.

Preference for Flexibility is the key axiom of Kreps (1979). In models such as Gul and Pesendorfer (2001) or Strotz (1955), this axiom characterizes the case of no conflict ($u \approx v$). Here, Preference for Flexibility permits the opposite case of total conflict. Intuitively, this



(a) Receiver's value

(b) Sender's value

Figure 4: Illustration of Proposition 1(iii) when $u \not\approx v$ and $u \not\approx -v$. The menu consists of three acts and Sender is indifferent between two of them regardless of beliefs. When no public information is provided, Sender chooses information yielding posteriors at $\hat{\mu}'$ and $\hat{\mu}_1 = 1$; for Receiver, the corresponding value is given by the upper red dot. If public information generating posteriors at $\hat{\mu}'$ and $\hat{\mu}''$ is provided, Sender chooses not to provide any additional information: doing so can only decrease payoffs (strictly so at posterior $\hat{\mu}'$). Consequently, Receiver's payoff decreases.

is so because temptation influences choice only via information provision. When $u \approx -v$, Sender chooses no-information at all menus; this reduces Receiver's behavior to standard expected-utility maximization, which satisfies Preference for Flexibility.

Preference for Information requires that public information never harms Receiver. Note that since $A \subseteq \sigma A$, this condition is implied by Preference for Flexibility; Proposition 1 establishes that for Persuasion Representations, it is in fact equivalent to Preference for Flexibility. The condition is satisfied by the $u \approx -v$ case since this reduces the representation to expected-utility maximization and, by Blackwell (1951,1953), the value of every decision problem increases with the availability of information. The $u \approx v$ case also satisfies Preference for Information since the representation reduces to $\overline{U}$ (Receiver's value under perfect information) making $\sigma A \sim A$ for all A. It is less obvious that menus A and experiments σ satisfying $A \succ \sigma A$ exist when $u \not\approx v$ and $u \not\approx -v$. As illustrated in Figure 4, the idea is that Sender may choose to provide nontrivial information $\hat{\sigma}$ at A but, under some public σ less-valuable to Receiver than $\hat{\sigma}$, be unwilling to provide any information.[25]

The fact that public information can harm Receiver via its effect on Sender's incentives has parallels in game theory, where being informed need not result in better equilibrium

[25]The example in Figure 4 is non-generic in that it relies on Sender being indifferent between two acts at all beliefs. This is for expository convenience. One can construct similar examples that do not involve such indifference, but it is more tedious to verify that they satisfy the desired properties.

outcomes for the informed player. Proposition 1 demonstrates that, for Receiver, such concerns vanish only in extreme cases and that public information is never harmful if and only if flexibility is never harmful.

Conditions (ii) and (iv) provide choice-correspondence analogues of Preference for Flexibility and Preference for Information, respectively. Condition (ii), Sen's α (Sen, 1971), is a basic property of rational choice; since the menus referenced by the axiom are $\subseteq$-comparable, the axiom constrains choice patterns that can arise under increased flexibility. Condition (iv) weakens Sen's α by requiring the original axiom to hold only when the increased flexibility is due to the availability of public information; nonetheless, the Proposition establishes that it is equivalent to Sen's α in Persuasion Representations. It is important to note that while Sen's α is equivalent to Preference for Flexibility in terms of the impact on u and v, Persuasion Representations of $\succsim$ involve sophistication (correct forecasting of future behavior) while representations for c do not. In this sense, Sen's α is a more robust test of the edge cases $u \approx \pm v$ than Preference for Flexibility; similarly, Informational Sen's α is more robust than Preference for Information.

Simple axioms characterize the $u \approx v$ and $u \approx -v$ cases. For any menu A and state ω, let $A_\omega := \{f_\omega : f \in A\}$. An experiment σ is **interior** if $0 < s_\omega < 1$ for all $s \in \sigma$ and $\omega \in \Omega$.

Proposition 2. *Suppose (μ, u, v) represents $\succsim$ and c. Then $u \approx v$ if and only if any of the following conditions hold:*

(i) $\succsim$ satisfies Preference for Statewise Flexibility: $A_\omega \supseteq B_\omega$ for all ω implies $A \succsim B$.

(ii) c is monotone in statewise flexibility: $A_\omega \supseteq B_\omega$ for all ω implies $c(A) \subseteq c(A \cup B)$.

(iii) $\succsim$ is indifferent to information: for all A and σ, $A \sim \sigma A$.

(iv) c is invariant to imperfect information: for all A and interior σ, $c(\sigma A) = c(A)$.

Proposition 2 characterizes the $u \approx v$ case in terms Receiver's preferences for, and choices under, increased flexibility or information. Condition (i) requires Receiver to prefer menus that offer more flexibility (more potential lotteries) in each state; this condition implies, but is not equivalent to, Preference for Flexibility. The analogous requirement for c, condition (ii), is that any act chosen at A is also chosen after expanding A in a way that does not expand the set of possible lotteries in any state. Intuitively, Sender chooses perfect information when $u \approx v$, so Receiver's behavior is not affected by additional options unless they expand the set of possible lotteries in some state. Conditions (iii) and (iv) capture the idea that public information has no impact on Receiver's welfare and choices when $u \approx v$; again, this holds because Sender chooses perfect information when there is no conflict with Receiver.

Proposition 3. *Suppose (μ, u, v) represents $\succsim$ and c. Then $u \approx -v$ if and only if any of the following conditions hold:*

(i) $\succsim$ is Independent of Irrelevant Alternatives: $A \succsim B$ implies $A \sim A \cup B$.

(ii) c satisfies WARP: $c(A) \cap B \neq \emptyset$ implies $c(B) \cap A \subseteq c(A)$.

(iii) $\succsim$ satisfies Preference for Information and $\sigma A \succ A$ for some A and σ.

(iv) c satisfies Informational Sen's α and $c(\sigma A) \neq c(A)$ for some A and interior σ.

Parts (i) and (ii) of Proposition 3 establish that familiar axioms—IIA for $\succsim$, WARP for c—characterize the $u \approx -v$ case. Traditionally, these axioms ensure $\succsim$ and c can be represented by maximization of a menu-independent utility function. In Persuasion Representations, this is consistent with $u \approx -v$ because Sender's resulting choice of no-information at all menus reduces Receiver's behavior to standard expected utility maximization. Parts (iii) and (iv) characterize $u \approx -v$ in terms of behavior under public information. In particular, satisfying Preference for Information or Informational Sen's α non-trivially is necessary and sufficient for $u \approx -v$ in Persuasion Representations. As noted in the discussion of Proposition 1, the characterizations involving c do not require sophistication on the part of Receiver, making them more robust tests; the same holds for Propositions 2 and 3.

For condition (iv) in each of Propositions 1–3, some care is needed in the interpretation of $c(\sigma A)$. Recall that σA merely *simulates* an environment where the choice set is A and Receiver observes a signal generated by σ in addition to that provided by Sender. To understand the difference, consider Example 2 above with public information σ^* (perfect information). There, Receiver chooses (f_1, g_2) from σA. Intuitively, perfect information leads Receiver to choose f in state 1 and g in state 2. Thus, both f and g would be chosen from A with perfect public information, but only (f_1, g_2) is chosen from the menu σA.

To translate $c(\sigma A)$ into a statement about choices from A under public information σ, two additional definitions are needed. For any A and σ, let $c_A(\sigma A) := \{f \in A : \exists \sum_{s \in \sigma} s f^s \in c(\sigma A)$ and $s \in \sigma$ such that $f^s = f\}$ denote the **projection** of $c(\sigma A)$ to A. For an experiment σ, let $\text{supp}(\sigma) := \{\frac{s}{\|s\|} : s \in \sigma\}$ denote its **support**. With this notation in place, the above characterizations involving choice data under public information can be re-formulated as:

Proposition 4. *Suppose (μ, u, v) represents c. Then:*

(i) $u \approx v$ *or* $u \approx -v$ *if and only if* c *is expansive in signals: supp*$(\sigma) \subseteq$ *supp*(σ') *implies* $c_A(\sigma A) \subseteq c_A(\sigma' A)$ *for all* A.

(ii) $u \approx v$ *if and only if, for all* A *and* σ, $c(A) = c_A(\sigma A)$.

(iii) $u \approx -v$ *if and only if* c *is expansive in signals and* $c(A) \neq c_A(\sigma A)$ *for some* A *and* σ.

Proposition 4 characterizes the extreme cases of Sender-Receiver conflict in terms of choices from A (not σA) under public information σ. Part (i) states there is either no conflict or total conflict if and only if expanding the set of posteriors induced by the public structure results in a larger set of acts being chosen from A. Parts (ii) and (iii) differ in whether this expansiveness property holds trivially or non-trivially: there is no conflict if and only if choices are unresponsive to public information, and there is total conflict if choice data is non-trivially expansive in public information.

4.2 Measures of Conflict

The characterizations in the previous section establish key relationships between Receiver's ex-ante value of (and ex-post choice under) flexibility and public information but are limited to extreme cases ($u \approx v$ or $u \approx -v$) regarding the conflict between the agents. This section develops finer comparisons between Sender and Receiver.

Definition 5. Let $u, v, \dot{v}$ be utility indices. Then $\dot{v}$ is **more u-aligned than** v (and v is **less u-aligned than** $\dot{v}$) if either $u \approx -v$ or $\dot{v} \approx \alpha u + (1-\alpha)v$ for some $\alpha \in [0, 1]$.

Definition 5 is the key definition of Ahn et al. (2019). Intuitively, a mixture $\dot{v} \approx \alpha u + (1-\alpha)v$ more closely resembles u than v does. So, holding u fixed, a Sender with index $\dot{v}$ "disagrees less" with Receiver than a Sender with index v, softening the conflict between the agents.

The aim of this section is to characterize more- or less-aligned utilities in terms of Receiver's value of, and choice under, increased flexibility or public information. To begin, consider the following comparative notions for menu preferences:

Definition 6. Let $\succsim$ and $\dot{\succsim}$ denote preferences on $\mathcal{A}$.

(i) $\dot{\succsim}$ **values flexibility more than** $\succsim$ if, for all $A \supseteq B$, $A \succ B$ implies $A \mathrel{\dot{\succ}} B$.

(ii) $\dot{\succsim}$ **values information more than** $\succsim$ if, for all A and σ, $\sigma A \succ A$ implies $\sigma A \mathrel{\dot{\succ}} A$.

The idea of Definition 6 is that one agent (represented by $\dot{\succsim}$) values flexibility more than another (represented by $\succsim$) if there are more instances where he strictly prefers an expanded

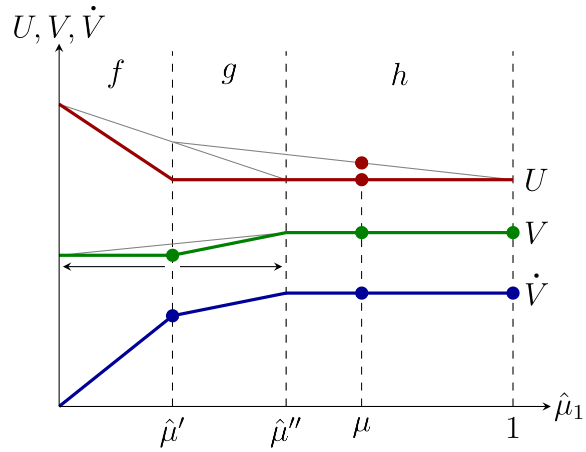


Figure 5: Illustration of Proposition 5(i). When no public information is provided, both Sender preferences (green and blue curves) result in no additional information provision; consequently, Receiver's payoff is the lower of the two red dots. Consider a public structure generating posteriors at $\hat{\mu}'$ and $\hat{\mu}_1 = 1$. The more-aligned Sender utility (green curve) compels Sender to provide additional information, yielding posteriors at $\hat{\mu}_1 = 0, \hat{\mu}''$, and $\hat{\mu}_1 = 1$, increasing Receiver's payoff to the higher red dot. However, the less-aligned Sender utility (blue curve) results in no additional information provision. Thus, less-aligned preferences do not increase Receiver's value of public information.

option set. Similarly, he values information more than the other agent if there are more instances where he strictly benefits from public information. Since $\sigma A \supseteq A$, an agent who values flexibility more than another necessarily values information more than the other.

Proposition 5. *Let* (μ, u, v) *and* $(\mu, u, \dot{v})$ *represent* $\succsim$ *and* $\dot{\succsim}$*, respectively.*

(i) If $\dot{\succsim}$ *values information more than* $\succsim$*, then* $\dot{v}$ *is less* u*-aligned than* v *is. However, the converse does not hold.*

(ii) If $\dot{\succsim}$ *values flexibility more than* $\succsim$*, then* $\dot{v}$ *is more* u*-aligned than* v *is and* $\dot{\succsim}$ *values information more than* $\succsim$ *does; consequently,* $\dot{v} \approx v$*.*

Part (i) of Proposition 5 states that if Receiver's value of public information increases, then there is greater conflict between the agents. Intuitively, the change in value reflects Receiver's expectation that Sender provides less information when there is greater conflict. However, the converse does not hold: less-aligned preferences do not guarantee that Receiver values information more in the sense of Definition 6. Figure 5 provides an example where $v = \frac{1}{2}u + \frac{1}{2}\dot{v}$. Proposition 6 below provides a finer condition that fully characterizes the degree of preference alignment in terms of Receiver's value of public information.

Part (ii) of Proposition 5 establishes that if Receiver's value of flexibility increases, then utilities become more aligned. However, as noted above, increased value of flexibility implies increased value of information; by part (i), then, utilities also become *less* aligned. Since $\dot{v}$ is both more and less u-aligned than v, it follows that $\dot{v} \approx v$. Thus, Definition 6(i) is too strong to characterize finer changes to the degree of conflict. The next result provides a remedy.

Proposition 6. *Let (μ, u, v) represent $\succsim$ and $(\mu, u, \dot{v})$ represent $\dot{\succsim}$. Then $\dot{v}$ is less u-aligned than v is if and only if any of the following conditions hold:*

(i) If $f \succsim A$, then $f \mathrel{\dot{\succsim}} A$.

(ii) If $f \sim \sigma A \succ A$ and $f \in \sigma A$, then $\sigma A \mathrel{\dot{\succ}} A$.

Proposition 6 modifies Definition 6 to provide full characterizations of utility alignment in Persuasion Representations. Part (i) establishes that utilities are less-aligned if there are more instances where full commitment (to a specific act) is preferred to a given menu A.[26] This captures the intuition that increased conflict lowers Receiver's expected payoff at A, enlarging the set of commitment options f that are preferred to playing the game with the full set A. Put differently, the result states that the concept of more-aligned utility fully characterizes the comparative statics of Receiver's welfare in persuasion games: his expected payoff increases at all menus if and only if utilities become more-aligned.

Part (ii) refines Definition 6(ii) to require that $\sigma A \mathrel{\dot{\succ}} A$ if $\sigma A \succ A$ *and* $f \sim \sigma A$ for some $f \in \sigma A$. In a Persuasion Representation, the latter requirement means f is prior-optimal at σA, indicating Sender does not provide *helpful* additional information beyond the public structure σ.[27] In this sense, the public structure is "binding"; for a less u-aligned index $\dot{v}$, σ is also binding: $f \mathrel{\dot{\sim}} \sigma A$ if $\sigma A \mathrel{\dot{\succ}} A$. Thus, part (ii) states there is greater conflict if and only if, for any A, there is a larger set of binding information structures that benefit Receiver.

To conclude this section, the next result characterizes the comparative statics in terms of choice data c.

Proposition 7. *Let (μ, u, v) represent c and $(\mu, u, \dot{v})$ represent $\dot{c}$. The following are equivalent:*

(i) $\dot{v}$ is less u-aligned than v is.

(ii) If $c(A) = f$, then $\dot{c}(A) = f$.

[26] Although f is not required to be an element of A, the characterization holds even if one imposes this restriction.

[27] Sender may be providing non-trivial information at σA to increase his own payoff, but $\sigma A \sim f \in \sigma A$ indicates Receiver does not benefit; eg, Receiver's prior-optimal act remains optimal at every posterior generated by Sender, so Receiver does not benefit but Sender may due to tie-breaking selections.

(iii) If $c(\sigma A) = f$*, then* $\dot{c}(\sigma A) = f$*.*

To understand Proposition 7, observe that $c(A) = f$ implies f is prior-optimal for Receiver at A. This means Sender discloses no information at A, indicating substantial disagreement regarding the value of outcomes generated by acts in A. In line with this intuition, the proposition states there is greater conflict between the agents if there are more menus where Receiver chooses only the prior-optimal act. For (iii), the logic (and proof) is almost identical but the result provides a different way of eliciting the relationship: given A, public information σ is binding (similarly to the discussion above) if it makes Sender choose not to disclose any additional information. The proposition thus states that preferences are less-aligned if and only if, for every A, the set of binding experiments expands.

4.3 Naivete and Sophistication

A central question—and the subject of ongoing research—concerns not just whether individuals experience temptation but whether they are sophisticated about (or aware of) their self-control problems. O'Donoghue and Rabin (1999,2001) contrast naivete and (partial or full) sophistication in settings of present-biased choice; since then, numerous studies have found support for partial sophistication in a variety of contexts.[28] In this section, I develop comparative notions of sophistication and naivete for Persuasion Representations, with a focus on two types of naivete: optimism and pessimism.

My approach builds on Ahn et al. (2019), who compare ex-ante preferences $\succsim$ *between* menus to ex-post choices *from* menus. The idea is that the function v revealed by $\succsim$ reflects Receiver's beliefs about Sender's preferences while that revealed by (say) λ reflects Sender's actual preferences; if these do not match, Receiver is (at most) partially sophisticated.

State-contingent random choice data λ is ideal for such comparisons. Given λ and A, let $f^{A,\lambda} := (\sum_{g \in A} \lambda^A_\omega(g) g_\omega)_{\omega \in \Omega}$; this is the state-contingent lottery generated Receiver's choices λ at A, and if λ has a Persuasion Representation it coincides with the induced act generated by composing Sender's chosen information structure with Receiver's signal-contingent choices.

Definition 7. Given λ, preferences $\succsim$ are **sophisticated** if $A \sim f^{A,\lambda}$ for all A.

The idea of Definition 7 is that λ, and therefore $f^{A,\lambda}$, reflects what actually happens at A while $\succsim$ captures Receiver's ex-ante expectations about behavior at A. If $A \sim f^{A,\lambda}$, there is no disconnect between expectations and reality: Receiver's ex-ante value of A coincides with

[28] There is a substantial literature investigating naivete and sophistication in lab and field settings; recent contributions include Cobb-Clark et al. (2024), Carrera et al. (2022), and Allcott et al. (2022).

the value of the induced act generated by subsequent behavior at A. Thus, sophistication (or lack thereof) is a property of $\succsim$ given some λ.[29]

Proposition 8. *Let (μ, u, v') represent $\succsim$ and (μ, u, v) represent λ. Then $\succsim$ is sophisticated if and only if $v' \approx v$.*

Proposition 8 verifies the intuition that a sophisticated Receiver holds correct beliefs about Sender's utility function: when $\succsim$ and λ have Persuasion Representations with common Receiver parameters (μ, u) but potentially different Sender utility functions v' and v, respectively, Receiver is sophisticated in the sense of Definition 7 if and only if $v' \approx v$. The remainder of this section studies the following two natural departures from sophistication.

Definition 8. Given λ, preferences $\succsim$ are **optimistic** if $A \succsim f^{A,\lambda}$ for all A. If instead $f^{A,\lambda} \succsim A$ for all A, preferences $\succsim$ are **pessimistic**.

Optimistic Receivers err only in an optimistic direction: if their ex-ante belief regarding the value of A differs from that of $f^{A,\lambda}$, it is because they expect a better outcome than $f^{A,\lambda}$. Similarly, pessimistic Receivers err only in the opposite direction. The next result provides a simple parametric representation of optimism and pessimism in persuasion models.

Proposition 9. *Let (μ, u, v) and (μ, u, v') represent λ and $\succsim$, respectively. Then:*

(i) $\succsim$ is optimistic if and only if v' is more u-aligned than v is.

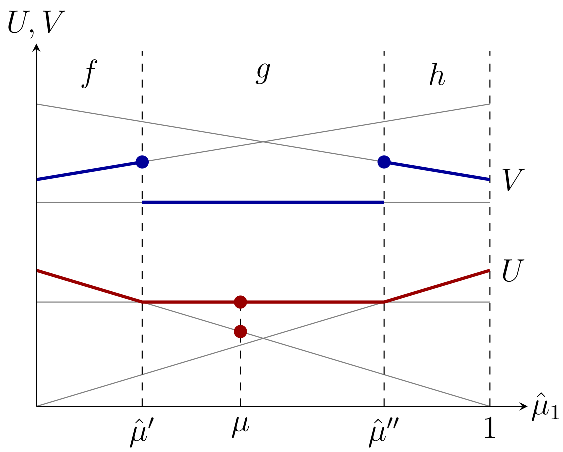
(ii) $\succsim$ is pessimistic if and only if v' is less u-aligned than v is.

Proposition 9 formalizes the intuition that optimism and pessimism correspond to Receiver's beliefs v' about Sender's utility being more or less u-aligned, respectively, than the true utility v. Thus, the requirements of Definition 8 characterize tight conditions on parameters in Persuasion Representations.

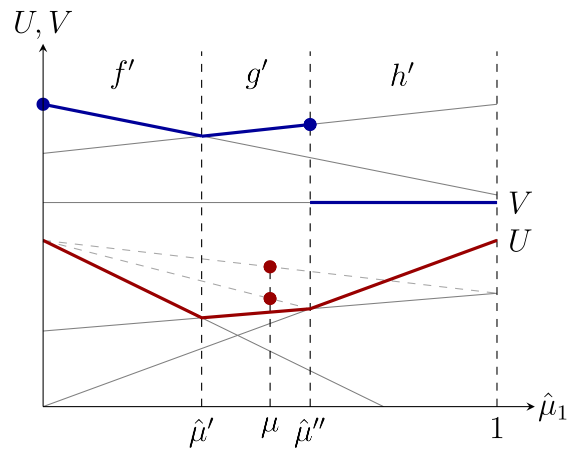
Proposition 10. *Let (μ, u, v) represent λ while (μ, u, v') and $(\mu, u, \dot{v})$ represent $\succsim'$ and $\dot{\succsim}$, respectively. Suppose $\succsim'$ and $\dot{\succsim}$ are optimistic. Then $\dot{v}$ is more u-aligned than v is (that is, $\dot{\succsim}$ is more optimistic than $\succsim'$) if and only if $A \succ' f^{A,\lambda}$ implies $A \mathrel{\dot{\succ}} f^{A,\lambda}$.*

Proposition 10 states that, conditional on being optimistic, increased optimism manifests as more instances of A being strictly preferred to $f^{A,\lambda}$. A symmetric result holds for pessimistic

[29] Ahn et al. (2019) define sophistication as $A \sim c(A)$ for all A. As explained below, this definition does not capture the desired behavior for Persuasion Representations.



(a) $A \succ c(A)$



(b) $c(B) \succ B$

Figure 6: In (a), Sender strongly dislikes g (Receiver's prior-optimal act in $A = \{f, g, h\}$), prompting information provision with posteriors at $\hat{\mu}'$ and $\hat{\mu}''$ and choices $c(A) = \{f, h\}$; Receiver's payoff is the upper red dot. Deleting g from A makes Sender's value function concave, resulting in no information provision and lowering Receiver's payoff to the lower red dot. In (b), Sender's preferences are mostly aligned with Receiver's but Sender strongly dislikes h' (Receiver's preferred act in state 1). Sender chooses information generating posteriors at $\hat{\mu}_1 = 0$ and $\hat{\mu}''$, yielding $c(B) = \{f', g'\}$ and Receiver payoffs at the lower red dot. Deleting h' from B makes Sender's value function convex, resulting in perfect information and raising Receiver's payoff to the upper red dot.

agents. This can be combined with the results of section 4.2 to establish, for example, that increased pessimism means greater value of public information (more instances where σA is preferred to A) and greater value of full commitment.

To conclude this section, it is worth noting that the definition of sophistication employed by Ahn et al. (2019), $A \sim c(A)$ for all A, does not capture the desired notion of sophistication for Persuasion Representations. Specifically, if (μ, u, v') represents $\succsim$ and (μ, u, v) represents c, then $v \approx v'$ does not imply $A \sim c(A)$ for all A. In other words, Receiver's welfare is affected by unchosen alternatives. Not only can $A \sim c(A)$ fail, but the preference can go in either direction: if $v \approx v'$, $u \not\approx v$ and $u \not\approx -v$, there are menus A, B such that $A \succ c(A)$ and $c(B) \succ B$. Figure 6 provides examples of such A and B.

5 Conclusion

This paper has developed a decision-theoretic analogue of the Bayesian Persuasion model in terms of Receiver's choices, preferences and welfare. The results establish that persuasion interactions can be understood entirely from Receiver's perspective: his behavior reveals all parameter values and, thereby, Sender's (unobserved) choice of information. While Sender has commitment power in choosing information, the framework is sufficiently rich to compare and contrast two ways of leveling the playing field for Receiver: hard commitment and public information provision. This leads to comparative static characterizations that, alongside the identification results, answer new and basic questions about the persuasion framework.

The paper also develops an interpretation of the framework as an intrapersonal interaction. Viewed this way, the model highlights the role of unobserved information acquisition for agents who experience temptation or other competing motives and must justify choices as consistent with their "true" objective. The identification results thus establish that standard choice data can reveal such motives as well as the endogenous, selective attention they generate. My approach introduces elements of the popular self-signaling paradigm into a decision-theoretic framework, enabling general comparisons between this and familiar decision-theoretic models of intrapersonal conflict. Both the foundations (see the Supplementary Appendix) and comparative statics are quite different when motivated attention is in play—important considerations for theoretical and empirical research involving temptation, demand for commitment, the value of information, sophistication/naivete, or justifiable choice.

A Proofs for Section 3

A.1 Proof of Theorem 2

Suppose $\succsim$ has a Persuasion Representation. It follows that the restriction of $\succsim$ to singleton menus is represented by subjective expected utility for some parameters (μ, u). By standard uniqueness arguments for the Anscombe-Aumann expected utility model, μ is unique and u is unique up to positive affine transformation.

To establish uniqueness of v, consider pq-bets $A_{pq} = \{pEq, qEp\}$ where p is interior and $u(p) > u(q)$. If $v(p) \geq v(q)$, Sender chooses perfect information and therefore $p \sim A_{pq}$. If $v(p) \leq v(q)$, Sender chooses e (no information) and $p \succ A_{pq}$. Therefore, the set $\{q : p \sim A_{pq} \succ q\}$ coincides with $\{q : u(p) > u(q) \text{ and } v(p) \geq v(q)\}$. Since p is interior, this set is a region in ΔX bounded by two planes: the indifference planes through p for u and v. Since u has been identified in the first step above, this reveals the indifference plane for v through p. The direction of increasing utility for v is also revealed by definition of the latter set. Thus, v is identified up to positive affine transformation.

A.2 Proof of Theorems 3 & 4

First, suppose a correspondence c has a Persuasion Representation. To identify u, consider menus $A = \{p, q\}$ of constant acts; for such menus, we have $p \in c(\{p, q\}) \Leftrightarrow u(p) > u(q)$ or $[u(p) = u(q)$ and $v(p) \geq v(q)]$. Consequently, $u(p) > u(q)$ if and only if there is a neighborhood around q such that $p \in c(\{p, q'\})$ for all q' in the neighborhood. Given p, then, the set of such q reveals the strict lower contour set of u through p, thereby revealing u up to positive affine transformation.

To identify v, consider pq-**menus** A where $u(p) > u(q)$; this means that for every $f \in A$, there is a signal s^f such that $f_\omega = s^f_\omega p + (1 - s^f_\omega)q$. Observe that if $v(p) \geq v(q)$, Sender chooses perfect information at A and so $c(A) = \overline{c}(A) := \bigcup_{\omega \in \Omega} \text{argmax}_{f \in A}\, u(f_\omega)$. If $v(p) < v(q)$, Sender chooses e (no information) and so $c(A) = \underline{c}(A) := \text{argmax}_{f \in A}\, U(f)$. Given $u(p) > u(q)$, it is straightforward to construct pq-menus A where $\overline{c}(A) \neq \underline{c}(A)$ and $c(A) = \overline{c}(A) \Leftrightarrow v(p) \geq v(q)$; for example, there is a pq-menu A where there is a unique prior-optimal act f^e (that is, $U(f^e) \geq U(g)$ for all $g \in A$) and, for each state ω, a unique act $f^\omega \in A$ such that $f^\omega_\omega = p$, making this act the unique optimal choice in state ω. Thus, $\overline{c}(A) = \{f^\omega : \omega \in \Omega\} \neq \{f^e\} = \underline{c}(A)$. Having identified u, then, such menus reveal the set $\{q : u(p) > u(q) \text{ and } v(p) \geq v(q)\}$; by the argument in the proof of Theorem 2 above, this reveals v up to positive affine transformation.

To identify μ, consider once again pq-menus A where $u(p) > u(q)$. Suppose $u \not\approx v$. Then there exists q such that $v(q) > v(p)$; consequently, there exist pq-menus A such that $c(A) = \underline{c}(A)$. Normalize $u(p) = 1$, $u(q) = 0$ and consider a pair of states $E = \{\omega, \omega'\}$ where $\omega \neq \omega'$. To pin down the ratio $\frac{\mu_\omega}{\mu_{\omega'}}$, consider a pq-menu A where all $f, f' \in A$ satisfy $f_{\hat{\omega}} = f'_{\hat{\omega}}$ for all $\hat{\omega} \notin E$ (that is, there exists an act $h = s^h p + (1 - s^h)q$ such that, for all $f \in A$, $f = fEh$). Since $c(A) = \underline{c}(A)$, it follows that $f \in c(A) \Leftrightarrow f \in \text{argmax}_{g \in A}\, s^g_\omega \mu_\omega + (1 - s^g_{\omega'})\mu_{\omega'}$. Thus, $f, g \in c(A) \Leftrightarrow (s^f_\omega - s^g_\omega)\mu_\omega = (s^g_{\omega'} - s^f_{\omega'})\mu_{\omega'}$. Appropriate choices of A thereby pin down $\frac{\mu_\omega}{\mu_{\omega'}}$. For example, letting $g = h = \frac{1}{2}p + \frac{1}{2}q$, one can elicit $s^f_\omega, s^f_{\omega'}$ such that $c(\{f, g\}) = \underline{c}(\{f, g\}) = \{f, g\}$, revealing $(s^f_\omega - \frac{1}{2})\mu_\omega = (\frac{1}{2} - s^f_{\omega'})\mu_{\omega'}$. Repeating this procedure for all pairs of states pins down all likelihood ratios and, therefore, pins down μ. This completes the proof of Theorem 4.

For Theorem 3, suppose ρ has a Persuasion Representation. Uniqueness of u and v follows from the argument for c, as does identification of μ if $u \not\approx v$. So, suppose $u \approx v$ and let $u(p) > u(q)$. Then $v(p) > v(q)$ and Sender chooses perfect information at all pq-menus A. In particular, for each state ω, consider a pq-bet $A = \{pEq, qEp\}$ where $E = \{\omega\}$. Since Sender chooses perfect information, it follows that $\rho^A(pEq) = \mu_\omega$, pinning down μ.

B Proofs for Section 4.1

Lemma 1. *Let u, v be non-constant utility indices such that $u \not\approx v$ and $u \not\approx -v$. For any pair of vectors $(u_1, \ldots, u_K), (v_1, \ldots, v_K) \in \mathbb{R}^K$, there is a set $\{p_1, \ldots, p_K\} \subseteq \Delta X$ and constants $A > 0$, $B, C \in \mathbb{R}$ such that, for all $k = 1, \ldots, K$, $u(p_k) = Au_k + B$ and $v(p_k) = Av_k + C$.*

Proof. Observe that ΔX can be identified with a subset of $\mathbb{R}^{N-1}$ (namely, the unit simplex in $\mathbb{R}^N$). Since u, v are non-constant linear functions on $\Delta X \subseteq \mathbb{R}^{N-1}$, their domains extend to all of $\mathbb{R}^{N-1}$ via linearity. For every k, the values u_k and v_k correspond to unique level sets (planes) of u and v in $\mathbb{R}^{N-1}$, respectively; since $u \not\approx v$ and $u \not\approx -v$ the normal vectors of these planes are linearly independent. Thus, for every k, there is a point $z^k \in \mathbb{R}^{N-1}$ such that $u(z^k) = u_k$ and $v(z^k) = v_k$. Pick a lottery p in the interior of ΔX. There is a scalar $\alpha \in (0, 1)$ sufficiently close to 1 such that, for all k, $\alpha p + (1 - \alpha)z^k \in \Delta X$; letting $p_k := \alpha p + (1 - \alpha)z^k$, $A := (1 - \alpha)$, $B := \alpha u(p)$ and $C := \alpha v(p)$ completes the proof. □

The significance of Lemma 1 is that it allows acts and menus to be constructed by selecting utility values for Receiver independently of the values for Sender when $u \not\approx v$ and $u \not\approx -v$. In particular, acts can be defined by specifying arbitrary profiles of utilities $(u_\omega)_{\omega \in \Omega}$ and $(v_\omega)_{\omega \in \Omega}$ (not necessarily in the range of u or v) and applying the lemma to obtain $f = (f_\omega)_{\omega \in \Omega}$ such that, for all ω, $u(f_\omega) = Au_\omega + B$ and $v(f_\omega) = Av_\omega + C$. More generally, menus are obtained by first specifying (for each act in the menu) utility profiles for Sender and Receiver and then applying the lemma to the full set of profiles—each act requires $|\Omega| = W$ lotteries, so a menu of M acts requires $K = MW$ lotteries. This way, the same constants $A > 0, B, C \in \mathbb{R}$ apply to every act, so in the resulting menu it is as if agents compare acts with the desired utility profiles. This simplifies the construction of many examples since value-function (or concavification) arguments only depend on the utility profiles, not the underlying lotteries.

B.1 Proof of Proposition 1

Proof of $u \approx v$ or $u \approx -v \Leftrightarrow$ (i). If $u \approx v$, then at every B Sender chooses perfect information; consequently, in every ω, both agents receive their most-preferred lottery in B_ω. If $B \subseteq A$, then the best outcome in each state can only improve under menu A. Thus, $A \succsim B$. If instead $u \approx -v$, then e (no information) is Sender-optimal at every menu A; consequently, Receiver chooses their prior-optimal act(s) from A. Thus, $\succsim$ reduces to a standard expected utility preference and therefore satisfies Preference for Flexibility.

For the converse, suppose $u \not\approx v$ and $u \not\approx -v$. Then, as is easily verified, there exist lotteries p, q, r such that $u(p) > u(r) > u(q)$ and $v(r) > v(p) > v(q)$. Let $E \neq \Omega$ be a nonempty subset of Ω and $B = \{pEq, qEp\}$. Since Sender and Receiver agree on the ranking of p and q, it follows that Sender chooses perfect information at menu B, yielding lottery p in every state. Thus, $U(B) = u(p)$.

Now let $A = \{pEq, qEp, r\}$. Clearly, $A \supseteq B$. We may choose r so that r is prior-optimal for Receiver in menu A; in particular, $U^e(pEq)$ and $U^e(qEp)$ belong to the open interval $(u(q), u(p))$ because $\mu(E) \in (0, 1)$. Thus, we may choose r near p (without reversing any inequalities above) so that $U^e(r) > \max\{U^e(pEq), U^e(qEp)\}$, making r prior-optimal for Receiver in menu A. Since $v(r) > v(p) > v(q)$, it follows that e (no information) is Sender-optimal at A (more generally, any Sender-optimal σ must yield Bayesian posteriors making r Receiver-optimal). Thus, outcome r is realized with probability 1, so that $U(A) = u(r) < u(p) = U(B)$, violating Preference for Flexibility. □

Proof of $u \approx v$ or $u \approx -v \Leftrightarrow$ (ii). First, suppose $u \approx v$ or $u \approx -v$. If $u \approx -v$, then (by the argument in part (i) above) c is rationalized by expected utility maximization with parameters (μ, u) and therefore satisfies Sen's α. If instead $u \approx v$, then (also by part (i) above) for all $\hat{A}$, we have $c(\hat{A}) = \{f \in \hat{A} : \exists \omega \text{ such that } u(f_\omega) \geq u(g_\omega)\ \forall g \in \hat{A}\}$. Let $f \in c(B) \cap A$ where $B \supseteq A$. Since $f \in c(B)$, there exists a state, say ω^*, such that $u(f_{\omega^*}) \geq u(g_{\omega^*})$ for all $g \in B \supseteq A$. Thus, $u(f_{\omega^*}) \geq u(g_{\omega^*})$ for all $g \in A$. Since $f \in A$, this implies $f \in c(A)$.

The converse is established by way of contradiction. So, suppose $u \not\approx v$ and $u \not\approx -v$. Choose an event E such that $0 < \mu(E) < 1$ and lotteries r, p, q, p', q' such that $u(p') > u(p') > u(r) > u(q') > u(q)$, $v(p) = v(q) > v(r) > v(p') = v(q')$, and $U^e(pEq) > u(r)$. Let $A = \{pEq, r\}$. Then e (no information) is Sender-optimal at A because pEq is prior-optimal for Receiver and $v(p) = v(q) > v(r)$; thus, $c(A) = \{pEq\}$. Now let $B = \{pEq, p'Eq', r\}$. Observe that, for Receiver, $p'Eq'$ dominates pEq. Thus, $p'Eq'$ is prior-optimal for Receiver; pEq is not chosen by Receiver at any signal; and r is chosen by Receiver at some signals because $0 < \mu(E) < 1$ and $u(p') > u(r) > u(q')$. Since $v(r) > v(p') = v(q')$, Sender selects an information structure where both $p'Eq'$ and r are chosen with positive probability; hence, $c(B) = \{p'Eq', r\}$. Thus, $r \in c(B) \cap A$ but $r \notin c(A)$, violating Sen's α. □

Proof of $u \approx v$ or $u \approx -v \Leftrightarrow$ (iii). First, suppose $u \approx v$ or $u \approx -v$. By part (i), $\succsim$ satisfies Preference for Flexibility. Since $\sigma A \supseteq A$, it follows that $\sigma A \succsim A$.

For the converse, suppose $u \not\approx v$ and $u \not\approx -v$. Consider first the case $|\Omega| = 2$. Since μ has full support, we may construct (by Lemma 1) a menu $A = \{f, g, h\}$ such that the value functions are of the form depicted by Figure 4 in the main text. In particular, Receiver finds f optimal on $[0, \hat{\mu}']$, g optimal on $[\hat{\mu}', \hat{\mu}'']$, and h optimal on $[\hat{\mu}'', 1]$, where the prior μ satisfies $\hat{\mu}' < \mu < \hat{\mu}''$. Sender is indifferent between g and h at all beliefs, prefers g (and h) to f on $[0, \hat{\mu}']$, and f to g and h on $[\hat{\mu}'', 1]$, with indifference between all three acts at $\hat{\mu}'$.

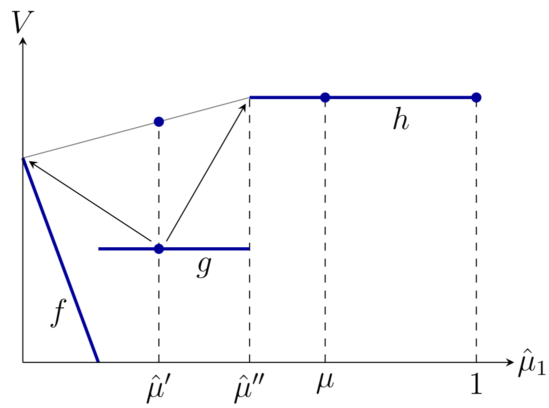


Figure 7: Illustration of Proposition 1(iv).

Let $\hat{\sigma}$ be an experiment such that, given μ, either posterior $\hat{\mu}'$ or $\hat{\mu} = 1$ is generated (such an experiment exists because $\hat{\mu}' < \mu < 1$). Observe that, at menu A, Sender finds both e and $\hat{\sigma}$ optimal because the concavification at $\hat{\mu}'$, μ, and $\hat{\mu} = 1$ coincides with Sender's value function. By Receiver-preferred tie breaking, Sender chooses $\hat{\sigma}$ as this yields the highest Receiver payoff among all Sender-optimal structures at A. Now consider a public structure σ that generates posteriors at $\hat{\mu}'$ and $\hat{\mu}''$. Again, Sender's values at these posteriors coincide with the concavification. However, Sender now prefers e because any non-trivial information structure creates a mean-preserving spread around both $\hat{\mu}'$ and $\hat{\mu}''$; in particular, any such spread around $\hat{\mu}'$ strictly lowers Sender's payoff. Thus, Sender chooses e and Receiver's value at σA is the prior-value of g and, thus, lower than the value at A.

For the general case $|\Omega| \geq 3$, let p be an arbitrary lottery, $E = \{\omega_1, \omega_2\}$ and consider the menu $A' := \{\hat{f}Ep : \hat{f} \in A\}$, where A is a menu of the form constructed above for the case $|\Omega| = 2$; this can be done because μ has full support. It is easy to see that $F(A') = \{\hat{f}Ep : \hat{f} \in F(A)\}$. Thus, at A, Sender selects an induced act f^*Ep, where $f^* \in F(A)$ is the induced act selected at A. Let σ' be an information structure generating posteriors corresponding to $\hat{\mu}'$ and $\hat{\mu}''$ in the 2-state construction above (conditional on E, the posteriors coincide with $\hat{\mu}'$ and $\hat{\mu}''$) and that leaves $\hat{\mu}'_\omega = \hat{\mu}''_\omega = \mu_\omega$ for all $\omega \notin E$. Then $F(\sigma' A') = \{\hat{f}Ep : \hat{f} \in \sigma A\}$, so Sender selects an induced act $g^*Ep \in F(\sigma' A')$ where $g^* \in F(\sigma A)$ is the induced act selected at σA. Thus, $A' \succ \sigma' A'$. □

Proof of $u \approx v$ or $u \approx -v \Leftrightarrow$ (iv). First, suppose $u \approx v$ or $u \approx -v$. By part (ii), c satisfies Sen's α. Since $\sigma A \supseteq A$, it follows that $c(\sigma A) \cap A \subseteq c(A)$, as desired.

For the converse, suppose $u \not\approx v$ and $u \not\approx -v$. Consider first the case $|\Omega| = 2$. Let $A = \{f, g, h\}$ be a menu such that Sender's value function is of the form depicted in Figure 5 (this is possible via Lemma 1 and the fact that μ has full support). Clearly, e is Sender-optimal at A and so $c(A) = \{h\}$. Consider the public structure σ generating posteriors at $\hat{\mu}'$ and $\hat{\mu} = 1$. Any additional information chosen by Sender generates a mean-preserving spread around $\hat{\mu}'$ but does not affect the point $\hat{\mu} = 1$. Thus, Sender chooses information $\hat{\sigma}$ to achieve the value of the concavification at $\hat{\mu}'$; this requires $\hat{\sigma}$ to generate posteriors at $\hat{\mu}''$ and at $\hat{\mu} = 0$. The latter posterior results in f being chosen. Thus, $f \in c(\sigma A) \cap A$ but $f \notin c(A)$. For the general case $|\Omega| \geq 3$, apply the same technique as the proof of part (ii) above to construct A' and $\sigma' A'$ such that $F(A')$ and $F(\sigma' A')$ are isomorphic to the sets $F(A)$ and $F(\sigma A)$ from the case $|\Omega| = 2$. □

B.2 Proof of Proposition 2

Proof of $u \approx v \Leftrightarrow$ (i). Suppose $u \approx v$. Then, for every menu B, perfect information is Sender-optimal because it yields u-maximal (hence v-maximal) lotteries from B_ω in every ω. If $B_\omega \subseteq A_\omega$ for all ω, then a maximal lottery in A_ω is maximal in $A_\omega \cup B_\omega$. Thus, $A \sim A \cup B$.

Conversely, suppose $u \not\approx v$. Then there are lotteries p, q such that $u(p) > u(q)$ and $v(q) > v(p)$. Let $\mu(E) \in (0, 1)$ and $A = \{pEq, qEp\}$. Sender and Receiver strictly disagree on the ranking of p and q, so Sender chooses e (no information) at A. Consequently, $U(A) \in (u(q), u(p))$ and $V(A) \in (v(p), v(q))$. Let $B = \{p\}$, so that $A \cup B = \{pEq, qEp, p\}$. Then p is a Receiver-optimal act in $A \cup B$ regardless of Sender's choice of information, implying $U(A \cup B) = u(p) > U(A)$; thus, $A \not\sim A \cup B$. Since $B_\omega = \{p\} \subseteq \{p, q\} = A_\omega$ for all ω, this contradicts Preference for Statewise Flexibility. □

Proof of $u \approx v \Leftrightarrow$ (iii). Suppose $u \approx v$. Then, for every A, perfect information is Sender-optimal and Receiver achieves a u-maximal lottery in A_ω for all ω. Let $\overline{u}_\omega$ denote Receiver's utility of any u-maximal lottery in A_ω. Observe that, for every σ and ω, Receiver again obtains utility $\overline{u}_\omega$ in state ω at menu σA (perfect information remains Sender-optimal and $A \subseteq \sigma A$). Thus, $A \sim \sigma A$.

Conversely, suppose $u \not\approx v$. Then there are lotteries p, q such that $u(p) > u(q)$ and $v(q) > v(p)$. Let $\mu(E) \in (0, 1)$ and $A = \{pEq, qEp\}$. As in the proof of $u \approx v \Leftrightarrow (i)$ above, this yields $U(A) \in (u(q), u(p))$. Consider $\sigma^* A$. Since $(p, \ldots, p) \in \sigma^* A$, we obtain $U(\sigma^* A) = u(p) > U(A)$, so that $A \not\sim \sigma^* A$. □

Proof of $u \approx v \Leftrightarrow$ (ii). Suppose $u \approx v$. Then perfect information is Sender-optimal; consequently, for all $\hat{A}$, we have

$$c(\hat{A}) = \{f \in \hat{A} : \exists \omega \text{ such that } u(f_\omega) \geq u(g_\omega) \ \forall g \in \hat{A}\}.$$

Suppose $B_\omega \subseteq A_\omega$ for all ω and let $f \in c(A)$. Then there is a state ω^* such that $u(f_{\omega^*}) \geq$

$u(g_{\omega^*})$ for all $g \in A$; thus, $u(f_{\omega^*}) \geq u(p)$ for all $p \in A_{\omega^*}$. Since $B_{\omega^*} \subseteq A_{\omega^*}$, it follows that $u(f_{\omega^*}) \geq u(q)$ for all $q \in B_{\omega^*}$. Then $u(f_{\omega^*}) \geq u(g_{\omega^*})$ for all $g \in A \cup B$, so that $f \in c(A \cup B)$.

For the converse, suppose $u \not\approx v$. As in the proof of $u \approx v \Leftrightarrow (i)$ above, the menus $A = \{pEq, qEp\}$ and $B = \{p\}$ satisfy $B_\omega \subseteq A_\omega$ for all ω but lead to $c(A \cup B) = \{p\}$, so that $c(A) \not\subseteq c(A \cup B)$. □

Proof of $u \approx v \Leftrightarrow$ (iv). Suppose $u \approx v$. As in the proof of $u \approx v \Leftrightarrow (ii)$ above, we have

$$c(A) = \{f \in A : \exists \omega \text{ such that } u(f_\omega) \geq u(g_\omega)\ \forall g \in A\}$$

for all A. Let σ denote an interior experiment. To see that $c(A) \subseteq c(\sigma A)$, first let $f \in c(A)$. Then $f \in \sigma A$ and $u(f_\omega) \geq u(p)$ for all $p \in A_\omega$. Thus, for every ω and $h \in \sigma A$, we have $u(f_\omega) \geq u(h_\omega)$ because h_ω is a mixture of lotteries in A_ω. Hence, $f \in c(\sigma A)$. To establish $c(\sigma A) \subseteq c(A)$, let $f \in c(\sigma A)$. This means there is a state ω^* such that $u(f_{\omega^*}) \geq u(h_{\omega^*})$ for all $h \in \sigma A \supseteq A$; thus, $u(f_{\omega^*}) \geq u(g_{\omega^*})$ for all $g \in A$. So, it will suffice to show that $f \in A$. Let $\omega \in \Omega$. A given lottery in $(\sigma A)_\omega \supseteq A_\omega$ is a mixture of lotteries in A_ω. If $f \in \sigma A \backslash A$ and the mixture f_ω assigns positive weight only to u-maximizers in A_ω, then σ cannot be interior: it must perfectly reveal state ω. Thus, $f \in A$.

For the converse, suppose $u \not\approx v$. Then there are lotteries p, q such that $u(p) > u(q)$ and $v(q) > v(p)$. Let $\mu(E) \in (0,1)$ and $A = \{pEq, qEp\}$. Let $\sigma^\varepsilon = [s^\varepsilon, e - s^\varepsilon]$ where $s^\varepsilon_\omega = 1 - \varepsilon$ for $\omega \in E$ and $s^\varepsilon_\omega = \varepsilon$ for $\omega \in \Omega \backslash E$. Then

$$\begin{aligned}\sigma^\varepsilon A &= \{pEq, qEp, s^\varepsilon pEq + (e - s^\varepsilon)qEp, s^\varepsilon qEp + (e - s^\varepsilon)pEq\} \\ &= \{pEq, qEp, (1-\varepsilon)p + \varepsilon q, \varepsilon p + (1-\varepsilon)q\}.\end{aligned}$$

Observe that e (no information) is Sender-optimal at $\sigma^\varepsilon A$. Thus, for $\varepsilon > 0$ sufficiently small, Receiver chooses $(1-\varepsilon)p + \varepsilon q$. Thus, $c(\sigma^\varepsilon A) \neq c(A)$. □

B.3 Proof of Proposition 3

Proof of $u \approx -v \Leftrightarrow (i)$. First, suppose $u \approx -v$. Then e (no information) is Sender-optimal at every menu A, so that Receiver chooses the prior-optimal act (according to u) from A. Thus, $\succsim$ reduces to a standard expected utility preference and, consequently, satisfies IIA.

Conversely, suppose $u \not\approx -v$. Then there are lotteries p, q such that $u(p) > u(q)$ and $v(p) > v(q)$. Let $A = \{pEq\}$ and $B = \{qEp\}$ where $1 > \mu(E) \geq \frac{1}{2}$. It follows that $U(A) = U^e(pEq) \geq U^e(qEp) = U(B)$, so that $A \succsim B$. However, at menu $A \cup B$, perfect information is Sender-optimal because it yields lottery p in every state. Thus, $U(A \cup B) = u(p) > U(A)$, so that $A \cup B \succ A$, contradicting IIA. □

Proof of $u \approx -v \Leftrightarrow (iii)$. First, suppose $u \approx -v$. By Proposition 1, $\succsim$ values information: $\sigma A \succsim A$ for all σ and A. Moreover, e (no information) is Sender-optimal at all menus A, so Receiver is indifferent between A and any U^e-optimal act $f \in A$. Consider a menu $A = \{pEq, qEp\}$ where $0 < \mu(E) < 1$ and $u(p) > u(q)$. Then $\sigma^* A \sim p \succ A$, where σ^* denotes perfect information.

For the converse, suppose $\sigma A \succsim A$ for all σ, A and that there exist σ, A such that $\sigma A \succ A$. By Proposition 1, either $u \approx v$ or $u \approx -v$. By Proposition 2, $u \not\approx v$ because $\succsim$ is not indifferent to information. Thus, $u \approx -v$. □

Proof of $u \approx -v \Leftrightarrow (ii)$. If $u \approx -v$, then Sender chooses e at every menu. Consequently, Receiver's choices are characterized by standard expected utility maximization and therefore satisfy WARP.

For the converse, suppose toward a contradiction that c satisfies WARP but $u \not\approx -v$. By WARP, there is a complete and transitive relation $\succsim$ such that, for all $\hat{A}$, $c(\hat{A}) = \{f \in \hat{A} : f \succsim g \ \forall g \in \hat{A}\}$. Since $u \not\approx -v$, there are lotteries p, q, r such that $u(p) > u(q) > u(r)$ and $v(p) > v(q) > v(r)$. Let E be a nonempty subset of Ω and let $f = rEp$, $g = r$, and $h = pEq$. In menu $A = \{f, g\}$, we have $c(A) = \{f, g\}$ because Sender's optimal information structure reveals whether the true state belongs to E or $\Omega \backslash E$ and both acts are chosen at states $\omega \in E$. Thus, the rationalizing preference satisfies $f \sim g$. In menu $B = \{g, h\}$, we have $c(B) = \{h\}$ because $u(p) > u(q) > u(r)$ implies act g is never chosen by Receiver. Thus, $h \succ g$. Finally, in menu $C = \{f, h\}$, we have $c(C) = \{f, h\}$ because Sender's optimal information structure reveals whether the true state belongs to E or $\Omega \backslash E$ and h is chosen for $\omega \in E$ while f is chosen for $\omega \notin E$. Thus, $f \sim h$. Combining these facts, we have $g \sim f \succ h \succ g$, a contradiction. □

Proof of $u \approx -v \Leftrightarrow (iv)$. First, suppose $u \approx -v$. By Proposition 1, c satisfies Informational Sen's α. Moreover, $c(A)$ consists of all U^e-optimal acts in A because e (no information) is Sender optimal in all menus. As in the proof of part (iv) of Proposition 2, let $0 < \mu(E) < 1$ and $\sigma^\varepsilon = [s^\varepsilon, e - s^\varepsilon]$ where $s^\varepsilon_\omega = 1 - \varepsilon$ for $\omega \in E$ and $s^\varepsilon_\omega = \varepsilon$ for $\omega \in \Omega \backslash E$; note that σ^ε is interior provided $0 < \varepsilon < 1$. Then, for any $A = \{pEq, qEp\}$ where $u(p) > u(q)$, we have $\sigma^\varepsilon A = \{pEq, qEp, (1-\varepsilon)p + \varepsilon q, \varepsilon p + (1-\varepsilon)q\}$. Observe that if $\varepsilon > 0$ is sufficiently close to 0, then $(1-\varepsilon)p + \varepsilon q$ is U^e-optimal in $\sigma^\varepsilon A$. Thus, $c(\sigma^\varepsilon A) = (1-\varepsilon)p + \varepsilon q \notin c(A)$. □

B.4 Proof of Proposition 4

(i) First, suppose $u \approx v$. Then perfect information is Sender-optimal at all menus B, yielding $c(B) = \{f \in B : \exists \omega \ u(f_\omega) \geq u(g_\omega) \ \forall g \in B\}$. For $B = \sigma A$, acts $f \in \sigma A$ are of the form $f = \sum_{s \in \sigma} s f^s$ ($f^s \in A$); consequently, $u(f_\omega) \geq u(g_\omega)$ for all $g \in \sigma A$ if and only if $u(\sum_{s\in\sigma} s_\omega f^s_\omega) \geq u(\sum_{s\in\sigma} s_\omega g^s_\omega)$ $\forall g^s \in A$ $(s \in \sigma)$. Thus, $f = \sum_{s\in\sigma} sf^s \in c(\sigma A)$ if and only if there is a state ω such that $u(f^s_\omega) \geq u(g_\omega)$ for all $s \in \sigma$ and $g \in A$. This implies $c_A(\sigma A) = c(A)$.

If $u \approx -v$, then e (no information) is Sender-optimal, yielding $c(B) = \{f \in B : U(f) \geq U(g)\ \forall g \in B\}$. For $B = \sigma A$, this implies $c(\sigma A) = \{f = \sum_{s\in\sigma} sf^s \in \sigma A : \forall s \in \sigma,\ U^s(f^s) \geq U^s(g)\ \forall g \in A\}$; thus, $\hat{f} \in c_A(\sigma A)$ if and only if there exists $s \in \sigma$ such that $U^s(\hat{f}) \geq U^s(g)$ for all $g \in A$. Suppose $\text{supp}(\sigma) \subseteq \text{supp}(\sigma')$ and let $\hat{f} \in c_A(\sigma A)$. Then there exists $s \in \sigma$ such that $U^s(\hat{f}) \geq U^s(g)$ for all $g \in A$. Since $\text{supp}(\sigma) \subseteq \text{supp}(\sigma')$, there exists $s' \in \sigma'$ such that s and s' yield the same Bayesian posterior. Thus, $U^{s'}(\hat{f}) \geq U^{s'}(g)$ for all $g \in A$ as well, so that $\hat{f} \in c_A(\sigma' A)$.

For the converse, suppose c is expansive in signals. Suppose toward a contradiction that $u \not\approx v$ and $u \not\approx -v$. Consider the case $|\Omega| = 2$ (this case extends to the general case via the arguments used in the proof of Proposition 1). Consider a menu $A = \{f, g, h\}$ with value function depicted in Figure 8; such a menu exists by Lemma 1. Sender is indifferent between g and h at all beliefs, strictly prefers f on $\hat{\mu}'_1 < \hat{\mu}_1 < 1$, and strictly prefers g (and h) on $0 < \hat{\mu}_1 < \hat{\mu}'$, with indifference between all three acts at $\hat{\mu}'$. Let $\sigma = e$ (so the only posterior it generates is the prior μ) and σ' be an information structure generating posteriors at μ, $\hat{\mu}'$, and $\hat{\mu}''$. At σ, Sender chooses (due to Receiver-preferred tie breaking) additional information so as to induce posteriors at $\hat{\mu}'$ and $\hat{\mu} = 1$. Consequently, $c_A(\sigma A) = \{f, g, h\}$ because h is chosen at $\hat{\mu} = 1$ while f and g are chosen at $\hat{\mu}'$. At $\sigma' A$, Sender chooses no additional information (each posterior yields a point on the concavification, and creating spread around $\hat{\mu}'$ strictly decreases Sender's payoff); consequently, $h \notin c_A(\sigma' A)$ because no posterior induced by σ' makes Receiver choose h. Thus, $\text{supp}(\sigma) \subseteq \text{supp}(\sigma')$ but $c_A(\sigma A) \not\subseteq c_A(\sigma' A)$.

(ii) First, suppose $u \approx v$. As demonstrated in the proof of (i), this implies $c(A) = c_A(\sigma A)$ for all A and σ. Conversely, suppose $u \not\approx v$. This implies there are lotteries p, q such that $u(p) > u(q)$ and $v(q) > v(p)$. Let $A = \{sp + (1-s)q, tp + (1-t)q\}$ be a pq-menu where Receiver has a unique prior-optimal act, say $sp + (1-s)q$. Let $\sigma = \sigma^*$ (perfect information). Then $c(A) = sp + (1-s)q$ but $c_A(\sigma A) = A$.

(iii) Suppose $u \approx -v$. By (i), c is expansive in signals and by (ii), $c(A) \neq c_A(\sigma A)$ for some A and σ (in particular, u and v are non-constant, so $u \approx -v$ implies $u \not\approx v$). Conversely, (i) and (ii) imply $u \approx -v$ if c is expansive in signals and $c(A) \neq c_A(\sigma A)$ for some A and σ.

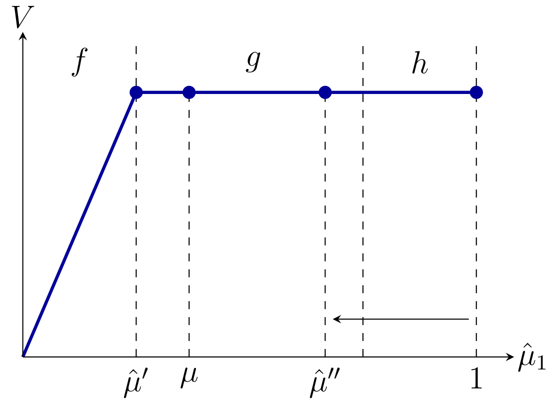


Figure 8: Illustration of Proposition 4(i).

C Proofs for Sections 4.2 and 4.3

Given utility indices u and v, a pair $\{p, q\}$ of lotteries is a (u, v)**-agreement pair** if either $[u(p) \geq u(q)$ and $v(p) \geq v(q)]$ or $[u(q) \geq u(p)$ and $v(q) \geq v(p)]$; otherwise, it is a **disagreement pair**.

Lemma 2. *Let $u, v, v' : X \to \mathbb{R}$ be non-constant utility indices such that $u \not\approx -v$ and $u \not\approx -v'$. Then every (u, v)-agreement pair is a (u, v')-agreement pair if and only if v' is more u-aligned than v.*

Proof. It is straightforward to show that if v' is more u-aligned than v, then every (u, v)-agreement pair is a (u, v')-agreement pair.

For the converse, let u, v, v' be non-constant utility indices such that $u \not\approx -v$ and $u \not\approx -v'$; interpret them as vectors $u = (u_1, \ldots, u_N)$, $v = (v_1, \ldots, x_N)$, and $v' = (v'_1, \ldots, v'_N)$ in $\mathbb{R}^N$ where $X = \{x_1, \ldots, x_N\}$.

Let $Z := \{\hat{u} \in \mathbb{R}^N : \sum_{n=1}^N \hat{u}_n = 0\}$; this is a hyperplane in $\mathbb{R}^N$ with normal vector $(1, \ldots, 1)$. Observe that for every utility index u', there exists $\hat{u} \in Z$ representing the same expected utility preferences; in particular, letting $B = \sum_{n=1}^N u'_n$, the index $\hat{u}$ where $\hat{u}_n = u'_n - B/N$ is a member of Z and a positive affine transformation of u'. Thus, every expected utility preference over ΔX is represented by an index in Z. Since scaling an index by a positive number does not affect expected utility preferences, we normalize the non-zero vectors in Z to length 1 by letting $C := \{\frac{\hat{u}}{\|\hat{u}\|} : 0 \neq \hat{u} \in Z\}$. Thus, every non-constant expected utility preference over ΔX is represented by a unique vector in C.[30]

Observe that if $u, v, v' \in C$, then $v' \approx \alpha u + (1 - \alpha)v$ for some $\alpha \in [0, 1]$ if and only if the vector v' belongs to the conic hull $C_{u,v} := \{\beta u + \gamma v : \beta \geq 0, \gamma \geq 0\} \subseteq Z$ of u and v.

[30]This method of normalizing the set of utility indices is due to Dekel and Lipman (2012).

Suppose $v' \not\approx \alpha u + (1-\alpha)v$ for all α. Since $C_{u,v} \subseteq Z$ and Z is isomorphic to $\mathbb{R}^{N-1}$, Farkas' lemma implies there is a vector $y \in Z$ such that $v' \cdot y < 0$ and $\hat{u} \cdot y \geq 0$ for all $\hat{u} \in C_{u,v}$; in particular, $u \cdot y \geq 0$ and $v \cdot y \geq 0$.

Let $q = (\frac{1}{N}, \ldots, \frac{1}{N}) \in \Delta X$ (that is, $q(x) = \frac{1}{N}$ for all $x \in X$). Letting $p = q + y$, we have $\sum_{n=1}^{N} p_n = \sum_{n=1}^{N} q_n + y_n = 1$ since $q \in \Delta X$ and $y \in Z$; if necessary, replace y with δy for sufficiently small $\delta > 0$ (this does not affect any inequalities above) to ensure $p_n = q_n + y_n \in [0,1]$ for all n. Thus, $p \in \Delta X$ as well. It follows that $0 \leq u \cdot y = u \cdot (p - q)$, so that $u \cdot p \geq u \cdot q$. Similarly, $v \cdot p \geq v \cdot q$, so $\{p, q\}$ is a (u,v)-agreement pair. However, $0 > v' \cdot y = v' \cdot (p-q)$, so that $v' \cdot p < v' \cdot q$. Since u, v and v' are linear functions and p, q satisfy $u \cdot p \geq u \cdot q$ and $v \cdot p \geq v \cdot q$, we may perturb p and q to ensure $u \cdot p > u \cdot q$ (and $v \cdot p \geq v \cdot q$) without violating $v' \cdot p < v' \cdot q$. Thus, $\{p, q\}$ is not a (u, v')-agreement pair. □

Lemma 3. *Suppose (μ, u, v) and (μ, u, v') represent $\succsim$ and $\succsim'$, respectively. Let $U, U' : \mathcal{A} \to \mathbb{R}$ denote the corresponding functions of the form (2) induced by these parameters. Suppose v' is more u-aligned than v. Then $U'(A) \geq U(A)$ for all A and $U'(B) = U(B)$ for all singleton menus B.*

Proof. We have $v' \approx \alpha u + (1-\alpha)v$ for some $\alpha \in [0,1]$. Since parameters (μ, u) are common to the representations of $\succsim$ and $\succsim'$, both involve the same set $F(A)$ of induced acts; that is, for all A,

$$U(A) = \max U^e(f) \text{ subject to } f \in \operatorname*{argmax}_{f' \in F(A)} V(f')$$

and

$$U'(A) = \max U^e(g) \text{ subject to } g \in \operatorname*{argmax}_{g' \in F(A)} V'(g'),$$

where $V, V' : F \to \mathbb{R}$ are expected utility with prior μ and indices v and v', respectively, and $U^e : F \to \mathbb{R}$ is expected utility with prior μ and index u. Clearly, then, $U'(B) = U(B)$ for all singleton menus B.

Let A be an arbitrary menu. The claim is trivial if $\alpha = 0$, so suppose $\alpha > 0$. Suppose $f \in \operatorname{argmax}_{f' \in F(A)} V(f')$ and $g \in \operatorname{argmax}_{g' \in F(A)} \alpha U^e(g') + (1-\alpha)V(g')$; thus, g' maximizes $V' \approx \alpha U^e + (1-\alpha)V$ on $F(A)$. Then

$$\begin{aligned} \alpha U^e(g) + (1-\alpha)V(g) &\geq \alpha U^e(f) + (1-\alpha)V(f) && \text{since } g \text{ maximizes } V' \text{ on } F(A) \\ &\geq \alpha U^e(f) + (1-\alpha)V(g) && \text{since } f \text{ maximizes } V \text{ on } F(A), \end{aligned}$$

which implies $\alpha U^e(g) \geq \alpha U^e(f)$ and thus $U^e(g) \geq U^e(f)$ since $\alpha > 0$. Since f and g are arbitrary maximizers of V and V', respectively, this implies $U'(A) \geq U(A)$. □

C.1 Proof of Proposition 5

(i) Let A be a pq-bet where $u(p) > u(q)$ and $v(p) < v(q)$. Then Sender chooses no information at A, so $\sigma^* A \succ A$. Since $\dot{\succsim}$ values information more than $\succsim$, this implies $\sigma^* A \mathrel{\dot{\succ}} A$, so $\dot{v}(p) < \dot{v}(q)$ as well. Thus, every disagreement pair for u, v is a disagreement pair for $u, \dot{v}$. By Lemma 2, this implies $\dot{v}$ is less u-aligned than v is.

To see that the converse does not hold, consider Figure 5. There, a menu $A = \{f, g, h\}$ is constructed so that Receiver chooses f on $0 \leq \hat{\mu} < \hat{\mu}'$, g on $\hat{\mu}' < \hat{\mu} < \hat{\mu}''$, and h on $\hat{\mu}'' < \hat{\mu} \leq 1$, making Sender's value function concave under $\dot{v}$ (by Lemma 1, such a menu exists). Consequently, Sender (under $\dot{v}$) chooses no information at A, so that Receiver's ex-ante value of A is given by the lower of the two red dots. Let σ be an experiment generating posteriors at $\hat{\mu}'$ and $\hat{\mu} = 1$. Again, concavity of Sender's value function under $\dot{v}$ implies no additional information is chosen; thus, $\sigma A \mathrel{\dot{\sim}} A$. Now consider Sender utility $v \approx \alpha u + (1 - \alpha)\dot{v}$. For appropriate values of α (in the figure, $\alpha = \frac{1}{2}$), no-information remains Sender optimal at A while non-trivial information is optimal at σA. In particular, Sender's value function under v is convex around $\hat{\mu}'$, so Sender selects information $\hat{\sigma}$ generating a mean-preserving spread of the point $\hat{\mu}'$ to $\hat{\mu} = 0$ and $\hat{\mu}''$, as indicated by the arrows in the figure (additional information has no impact on the other posterior, $\hat{\mu} = 1$, generated by σ since it is a degenerate distribution). This structure is Sender-optimal at A because it raises Sender's payoff at $\hat{\mu}'$ to the point on the concavification at $\hat{\mu}'$. This information structure raises Receiver's payoff conditional at $\hat{\mu}'$ as well which, in turn, raises his overall payoff at σA to the higher of the two red dots. Thus, $\dot{v}$ is less u-aligned than v is but $\dot{\succsim}$ does not value information more than $\succsim$ because $\sigma A \succ A$ while $\sigma A \mathrel{\dot{\sim}} A$.

(ii) Suppose $\dot{\succsim}$ values flexibility more than $\succsim$. Let p, q be an agreement pair for u, v and $A = \{pEq, qEp\}$ a pq-bet. If $u(p) = u(q)$, then p, q is automatically an agreement pair for $u, \dot{v}$. If instead $u(p) \neq u(q)$, suppose without loss that $u(p) > u(q)$. Then $A \succ \{pEq\}$ (where pEq is Receiver's prior-optimal act in A) and so $A \mathrel{\dot{\succ}} \{pEq\}$ as well. This implies the $\dot{\succ}$-Sender chooses perfect information at A, and so p, q is an agreement pair for $u, \dot{v}$. Thus, every u, v agreement pair is a $u, \dot{v}$ agreement pair, making $\dot{v}$ more u-aligned than v (Lemma 2). However, $\dot{\succsim}$ also values information more than $\succsim$ because $\sigma A \supseteq A$ for all A and σ. Thus, by (i), $\dot{v}$ is also less u-aligned than v. Consequently, $\dot{v} \approx v$.

C.2 Proof of Proposition 6

(i) Suppose $\dot{v}$ is less u-aligned than v is. Since the set of induced acts only depends on μ and u, it follows from Lemma 3 that Receiver's value of A under (μ, u, v) weakly exceeds that under $(\mu, u, \dot{v})$. Thus, $f \succsim A$ implies $f \mathrel{\dot{\succsim}} A$.

Conversely, suppose statement (i) holds and consider a pq-bet A where $u(p) > u(q)$. Let $f \in A$ denote $\dot{\succsim}$-Receiver's prior-optimal act. Then $A \mathrel{\dot{\succ}} f$ if and only if $\dot{v}(p) \geq \dot{v}(q)$. A similar result holds with v in place of $\dot{v}$ and $\succ$ in place of $\dot{\succ}$. Therefore, statement (i) (in contrapositive form) implies every agreement pair for $u, \dot{v}$ is an agreement pair for u, v. Lemma 2, then, implies v is more u-aligned than $\dot{v}$ is.

(ii) Suppose $\dot{v}$ is less u-aligned than v is and that $f \sim \sigma A \succ A$ for some $f \in \sigma A$. This implies f is prior-optimal for Receiver in σA and, thus, U-minimal in $F(\sigma A)$. Since $F(\sigma A)$ only depends on μ and u, Lemma 3 implies $\dot{\succsim}$-Sender selects a U-minimal act from $F(\sigma A)$ as well. Thus, $f \mathrel{\dot{\sim}} \sigma A \mathrel{\dot{\succ}} A$.

Conversely, suppose statement (ii) holds and consider a pq-bet A where $u(p) \neq u(q)$. If $f \sim \sigma^* A \succ A$, then u, v disagree on the ranking of p, q. Condition (ii) implies $\sigma^* A \mathrel{\dot{\succ}} A$, which means $u, \dot{v}$ disagree on the ranking as well. Thus, every disagreement pair for u, v is a disagreement pair for $u, \dot{v}$; by Lemma 2, then, $\dot{v}$ is less u-aligned than v is.

C.3 Proof of Proposition 7

First, suppose $\dot{v}$ is less u-aligned than v is. If $c(A) = f$, then f is prior-optimal for Receiver, so c-Sender must be choosing e (no information) at A. Since $\dot{v}$ is less u-aligned than v is, it follows from Lemma 3 that $\dot{c}$-Sender chooses e as well. Thus, $\dot{c}(A) = f$.

Conversely, consider a pq-bet A. If $c(A) = f$, the c-agents disagree on the ranking of p, q; by (ii), we have $\dot{c}(A) = f$ and so the $\dot{c}$-agents disagree as well. Thus, every disagreement pair for u, v is a disagreement pair for $u, \dot{v}$. By Lemma 2, then, $\dot{v}$ is less u-aligned than v.

These argument for (i) $\Rightarrow$ (ii) holds if one replaces A with σA for any σ, establishing (i) $\Rightarrow$ (iii). For (iii) $\Rightarrow$ (i), let A be a pq-bet and σ a non-perfect information structure. If $c(\sigma A) = f$, this implies p, q is a u, v disagreement pair. By (iii), $\dot{c}(\sigma A) = f$, implying p, q is a $u, \dot{v}$ disagreement pair. By Lemma 2, then, $\dot{v}$ is less u-aligned than v.

C.4 Proof of Proposition 8

Clearly, $\succsim$ is sophisticated if $v' \approx v$. Conversely, suppose $\succsim$ is sophisticated. There are three cases:

1. $u \approx v$. If $v' \not\approx v$, there exist p, q such that $u(p) > u(q)$ and $v'(p) < v'(q)$; a corresponding pq-bet A yields $p \succ A$ since a Sender with preferences v' chooses no information in this case. But sophistication together with $u \approx v$ implies $A \sim f_\lambda^A = p$ for all pq-bets A where $u(p) \geq u(q)$, a contradiction.

2. $u \approx -v$. If $v' \not\approx v$, there exist p, q such that $u(p) > u(q)$ and $v'(p) > v'(q)$; a corresponding pq-bet satisfies $p \sim A$ since a Sender with preferences v' chooses perfect information in this case. But sophistication together with $u \approx -v$ implies $p \succ f_\lambda^A \sim A$ for all pq-bets A where $u(p) > u(q)$, a contradiction.

3. $u \not\approx v$ *and* $u \not\approx -v$. Suppose $v' \not\approx v$. If $v' \approx u$ or $v' \approx -u$, the argument from case 1 or 2 applies. If $v' \not\approx u$ and $v' \not\approx -u$, there exist p, q such that $u(p) > u(q)$, $v(p) > v(q)$, and $v'(p) < v'(q)$; now the argument from case 1 applies.

C.5 Proof of Proposition 9

We prove statement (i) (the proof for (ii) is similar). First, suppose v' is more u-aligned than v is. Let $U' : \mathcal{A} \to \mathbb{R}$ denote Receiver's value (Persuasion Representation) of $\succsim$ with parameters (μ, u, v'), and $U : \mathcal{A} \to \mathbb{R}$ Receiver's value in a Persuasion Representation with parameters (μ, u, v). Then $U'(f) = U(f)$ for all f and, by Lemma 3, $U'(A) \geq U(A)$ for all A. Suppose $A \not\sim f_\lambda^A$. Then $U'(A) \neq U'(f_\lambda^A) = U(f_\lambda^A) = U(A)$; thus, $U'(A) > U(A) = U'(f_\lambda^A)$ and so $A \succ f_\lambda^A$.

For the converse, suppose $\succsim$ is optimistic. We show that every (u, v)-agreement pair is a (u, v')-agreement pair; the desired result then follows from Lemma 2. Suppose toward a contradiction that there exists a (u, v)-agreement pair $\{p, q\}$ that is a (u, v')-disagreement pair. Without loss of generality, suppose $u(p) > u(q)$; then $v(p) \geq v(q)$ and $v'(p) < v'(q)$. Let A be a pq-bet. Then λ-Sender chooses perfect information but $\succsim$-Receiver expects $\succsim$-Sender to choose no information; consequently, $A \not\sim f_\lambda^A = p$. Since $\succsim$ is optimistic, we have $A \succ f_\lambda^A = p$; but $p \succsim A$ because A is a pq-bet and $u(p) > u(q)$. Thus, $\{p, q\}$ is a (u, v')-agreement pair.

C.6 Proof of Proposition 10

Suppose $\dot{v}$ is more u-aligned than v'. Let $U, U', \dot{U}$ denote Receiver's values (utility representations of menu preferences) under parameters (μ, u, v), (μ, u, v'), and $(\mu, u, \dot{v})$, respectively. By Lemma 3, we have $\dot{U}(A) \geq U'(A) \geq U(A)$ for all A and $\dot{U}(f) = U'(f) = U(f)$ for all f. Suppose $A \succ' f_\lambda^A$. Then $U'(A) > U'(f_\lambda^A) = U(f_\lambda^A) = U(A)$, so $\dot{U}(A) \geq U'(A) > U(A) = U(f_\lambda^A) = \dot{U}(f_\lambda^A)$. Thus, $A \mathrel{\dot{\succ}} f_\lambda^A$.

For the converse, suppose $A \succ' f_\lambda^A$ implies $A \mathrel{\dot{\succ}} f_\lambda^A$. We show that every (u, v')-agreement pair is a $(u, \dot{v})$-agreement pair. So, let $\{p, q\}$ be a (u, v')-agreement pair; without loss of generality, assume $u(p) > u(q)$ and $v'(p) \geq v'(q)$. Let A be a pq-bet. There are two cases. First, suppose $\{p, q\}$ is a (u, v)-agreement pair. Since $\dot{\succsim}$ is optimistic, $\{p, q\}$ is also a $(u, \dot{v})$-agreement pair (see the proof of Proposition 9). Second, suppose instead that $\{p, q\}$ is a (u, v)-disagreement pair. Then $A \succ' f_\lambda^A$ because λ-Sender chooses no information but $\succsim'$-Receiver believes $\succsim'$-Sender chooses perfect information since $v'(p) \geq v'(q)$. By hypothesis, it follows that $A \mathrel{\dot{\succ}} f_\lambda^A$. Since A is a pq-bet and $\{p, q\}$ is a (u, v)-disagreement pair, this implies λ-Sender chooses no information but $\dot{\succsim}$-Receiver believes $\dot{\succsim}$-Sender chooses perfect information; thus, $\dot{v}(p) \geq \dot{v}(q)$, so that $\{p, q\}$ is a $(u, \dot{v})$-agreement pair.

References

Ahn, D. S., R. Iijima, Y. Le Yaouanq, and T. Sarver (2019). Behavioural characterizations of naivete for time-inconsistent preferences. *The Review of Economic Studies 86*(6), 2319–2355.

Ainslie, G. (1992). *Picoeconomics: The strategic interaction of successive motivational states within the person.* Cambridge University Press.

Allcott, H., M. Gentzkow, and L. Song (2022). Digital addiction. *American Economic Review 112*(7), 2424–2463.

Anscombe, F. J. and R. J. Aumann (1963). A definition of subjective probability. *The annals of mathematical statistics 34*(1), 199–205.

Azrieli, Y. and J. Rehbeck (2022). Marginal stochastic choice. *arXiv preprint arXiv:2208.08492*.

Bénabou, R. and J. Tirole (2002). Self-confidence and personal motivation. *The quarterly journal of economics 117*(3), 871–915.

Bénabou, R. and J. Tirole (2004). Willpower and personal rules. *Journal of Political Economy 112*(4), 848–886.

Bénabou, R. and J. Tirole (2006). Incentives and prosocial behavior. *American economic review 96*(5), 1652–1678.

Blackwell, D. (1951). Comparison of experiments. In *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, Volume 1, pp. 93–102.

Blackwell, D. (1953). Equivalent comparisons of experiments. *The annals of mathematical statistics 24*(2), 265–272.

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

Carrera, M., H. Royer, M. Stehr, J. Sydnor, and D. Taubinsky (2022). Who chooses commitment? evidence and welfare implications. *The Review of Economic Studies 89*(3), 1205–1244.

Carrillo, J. D. and T. Mariotti (2000). Strategic ignorance as a self-disciplining device. *The Review of Economic Studies 67*(3), 529–544.

Cherepanov, V., T. Feddersen, and A. Sandroni (2013). Rationalization. *Theoretical Economics 8*(3), 775–800.

Cobb-Clark, D. A., S. C. Dahmann, D. A. Kamhöfer, and H. Schildberg-Hörisch (2024). Sophistication about self-control. *Journal of Public Economics 238*, 105196.

Crawford, V. P. and J. Sobel (1982). Strategic information transmission. *Econometrica*, 1431–1451.

Curello, G. and L. Sinander (2022). The comparative statics of persuasion. *arXiv preprint arXiv:2204.07474*.

Dana, J., R. A. Weber, and J. X. Kuang (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory 33*(1), 67–80.

de Oliveira, H. (2018). Blackwell's informativeness theorem using diagrams. *Games and Economic Behavior 109*, 126–131.

Dekel, E. and B. L. Lipman (2012). Costly self-control and random self-indulgence. *Econometrica 80*(3), 1271–1302.

Dekel, E., B. L. Lipman, and A. Rustichini (2009). Temptation-driven preferences. *The Review of Economic Studies 76*(3), 937–971.

Dillenberger, D., J. S. Lleras, P. Sadowski, and N. Takeoka (2014). A theory of subjective learning. *Journal of Economic Theory 153*, 287–312.

Doval, L., R. Eilat, T. Liu, and Y. Zhou (2024). Revealed information. *arXiv preprint arXiv:2411.13293*.

Ellis, A. (2018). Foundations for optimal inattention. *Journal of Economic Theory 173*, 56–94.

Ergin, H. and T. Sarver (2010). A unique costly contemplation representation. *Econometrica 78*(4), 1285–1339.

Exley, C. L. (2016). Excusing selfishness in charitable giving: The role of risk. *The Review of Economic Studies 83*(2), 587–628.

Exley, C. L. and J. B. Kessler (2023). Information avoidance and image concerns. *The Economic Journal 133*(656), 3153–3168.

Exley, C. L. and J. B. Kessler (2024). Motivated errors. *American Economic Review 114*(4), 961–987.

Falmagne, J.-C. (1978). A representation theorem for finite random scale systems. *Journal of Mathematical Psychology 18*(1), 52–72.

Golman, R., D. Hagmann, and G. Loewenstein (2017). Information avoidance. *Journal of economic literature 55*(1), 96–135.

Grossman, Z. (2015). Self-signaling and social-signaling in giving. *Journal of Economic Behavior & Organization 117*, 26–39.

Grossman, Z. and J. J. Van Der Weele (2017). Self-image and willful ignorance in social decisions. *Journal of the European Economic Association 15*(1), 173–217.

Gul, F. and W. Pesendorfer (2001). Temptation and self-control. *Econometrica 69*(6), 1403–1435.

Gul, F. and W. Pesendorfer (2006). Random expected utility. *Econometrica 74*(1), 121–146.

Haisley, E. C. and R. A. Weber (2010). Self-serving interpretations of ambiguity in other-regarding behavior. *Games and economic behavior 68*(2), 614–625.

Jakobsen, A. M. (2021). An axiomatic model of persuasion. *Econometrica 89*(5), 2081–2116.

Kalai, G., A. Rubinstein, and R. Spiegler (2002). Rationalizing choice functions by multiple rationales. *Econometrica 70*(6), 2481–2488.

Kamenica, E. and M. Gentzkow (2011). Bayesian persuasion. *American Economic Review 101*(6), 2590–2615.

Kreps, D. M. (1979). A representation theorem for "preference for flexibility". *Econometrica: Journal of the Econometric Society*, 565–577.

Lehrer, E. and R. Teper (2011). Justifiable preferences. *Journal of Economic Theory 146*(2), 762–774.

Lipman, B. and W. Pesendorfer (2013). Temptation. In *Advances in economics and econometrics: Tenth World Congress*, Volume 1, pp. 243–288. Citeseer.

Lu, J. (2016). Random choice and private information. *Econometrica 84*(6), 1983–2027.

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

O'donoghue, T. and M. Rabin (2001). Choice and procrastination. *The Quarterly Journal of Economics 116*(1), 121–160.

Rehbeck, J. (2023). Revealed bayesian expected utility with limited data. *Journal of Economic Behavior & Organization 207*, 81–95.

Ridout, S. (2023). Distinguishing preferences from justifications. *Working paper*.

Robbett, A., H. Walsh, and P. H. Matthews (2024). Moral wiggle room and group favoritism among political partisans. *PNAS nexus 3*(10), pgae307.

Sarver, T. (2008). Anticipating regret: Why fewer options may be better. *Econometrica 76*(2), 263–305.

Schilbach, F. (2019). Alcohol and self-control: A field experiment in india. *American economic review 109*(4), 1290–1322.

Sen, A. K. (1971). Choice functions and revealed preference. *The Review of Economic Studies 38*(3), 307–317.

Smith, A. (1759). *The theory of moral sentiments.*

Strotz, R. H. (1955). Myopia and inconsistency in dynamic utility maximization. *The Review of Economic Studies 23*(3), 165–180.

Toussaert, S. (2018). Eliciting temptation and self-control through menu choices: A lab experiment. *Econometrica 86*(3), 859–889.

Wang, Z. (2022). Regret and information avoidance. *Available at SSRN 4269560*.