Ambiguous Persuasion*

Dorian Beauchêne† Jian Li‡ Ming Li§

October 2017
Preliminary and incomplete, do not cite or circulate.

Abstract

We study a persuasion game à la Kamenica and Gentzkow (2011) where players are ambiguity averse with maxmin expected utility (Gilboa and Schmeidler, 1989). With no prior ambiguity, Sender might choose to use ambiguous communication devices. The main result characterizes the value of optimal ambiguous persuasion, which is often higher than what is feasible under Bayesian persuasion. We characterize posterior beliefs that are potentially plausible with being generated by some ambiguous devices. One way to construct an optimal ambiguous communication device is by using synonyms, messages that lead to the same posteriors, in which Sender can hedge himself against ambiguity while inducing actions from Receiver that would not be possible under Bayesian communication. Two applications, including the well-known uniform-quadratic example, are considered. Our analysis provides a justification for how ambiguity may emerge endogenously in persuasion.

*We are grateful to Pierre Fleckinger, Françoise Forges, Sidartha Gordon, Yingni Guo, Sean Horan, Maxim Ivanov, Frédéric Koessler, Marie Laclau, Elliot Lipnowski, Sujoy Mukerji, Peter Norman, Eduardo Perez-Richet, Ludovic Renou, Joel Sobel, Jean-Marc Tallon, Marciano Siniscalchi, and Vassili Vergopoulos for their very helpful comments. We are also grateful to audiences at Stony Brook Game Theory Conference 2016, Midwest Theory conference 2016, 2017, Midwest Political Science Association Conference 2017, Canadian Economic Theory Conference 2017, Asian Summer Meeting of Econometric Society 2017, McMaster University, McGill University, University of Montreal, and Laval University for valuable inputs. This research has benefitted from the research programs "Investissements d'avenir" ANR-10-LABX-93 and "Jeux et ambiguïté" ANR-12-FRAL-0008-01 granted by the Agence Nationale de la Recherche (Beauchêne), "Scientific research, conflicts of interest, and disclosure policy" (SSHRC Insight Development 430-2016-00444, Ming Li), and "Communication avec un public averse à l'ambiguïté" (FRQSC 2018-NP-206978, Jian Li). Part of the research was conducted while Beauchêne, Dorian was writing his Ph.D. dissertation at Paris School of Economics, and while Li, Ming was visiting ISER at Osaka University, who is grateful for the hospitality.

†MAPP Economics. E-mail: dorianbeauchene@gmail.com
‡McGill University and CIREQ. Email: jian.li7@mcgill.ca.
§Concordia University and CIREQ. Email: ming.li@concordia.ca.

1 Introduction

"If I asked for a cup of coffee, someone would search for the double meaning."

"When I'm good, I'm very good. But when I'm bad I'm better."

-Mae West

"Wording should not be varied capriciously, because in general people assume that if someone uses two different words they are referring to two different things."

-Steve Pinker[1]

Ambiguity is present in many settings of persuasion. Countries often keep their foreign policy intentionally ambiguous, a prominent example being the United States government's stance on whether it will defend Republic of China (Taiwan) if it comes under attack by People's Republic of China (Mainland China). Credit ratings agencies provide credit ratings that are left very coarse, even though it is clearly in the interest of investors to have a more precise rating and market competition should in principle push these agencies to provide a more precise one. Manufacturers of brand-name drugs often emphasize the uncertainty about the effectiveness and safety of their generic competitors.[2] Finally, Alan Greenspan has taken pride in perfecting the art of "Fed-Speak," with which he "would catch myself in the middle of a sentence" and "continue on resolving the sentence in some obscure way which made it incomprehensible."[3]

In all the examples above, a sender who controls access to some information chooses how to communicate with a decision maker who is uninformed—a receiver. Sender chooses a communication strategy and commits to it. Such environments have been studied by Kamenica and Gentzkow (2011) in a "Bayesian persuasion" framework, where Sender persuades Receiver by selecting a pre-committed communication device (or "signal," in the terminology of Kamenica and Gentzkow 2011). In contrast to Kamenica and Gentzkow (2011), who assume that Sender and Receiver are both expected utility maximizers, we allow Sender and Receiver to be ambiguity averse and allow the communication device to be ambiguous. We investigate *when* and *how* Sender can benefit from using ambiguous communication devices.

[1]See Pinker (2015).

[2]Merck, one of the world's largest pharmaceutical companies, sponsored a dinner-and-talk event for health professionals in Sault Ste. Marie, Ontario in 2014, where the talk included themes like "Generic medications: are they really equal?" and "Do generics help or hinder patient care?" and was given by Dr. Peter J. Lin, a prominent Canadian health commentator and family physician, who had repeatedly questioned the benefits of generic drugs and the reliability of the Canadian government's approval procedure for generic drugs. See Blackwell (2014).

[3]See Leonard, Devin and Peter Coy, August 13-26, 2012, "*Alan Greenspan on His Fed Legacy and the Economy*," BusinessWeek: 65.

Going back to the example of a brand name drug producing pharmaceutical company (Sender/"he") who wants to persuade a physician (Receiver/"she") to refrain from prescribing the genetic competitor of one of his drugs. The pharmaceutical company could commission studies on the (in)effectiveness and (un)safety of the generic drug. If the physician were Bayesian, as Kamenica and Gentzkow (2011) assume, she would form a belief about the effectiveness and safety of the generic drug based on the results of the studies and trade it off against the extra expense of the brand-name drug. She will then make the decision on whether to prescribe the generic drug based on her belief. Under these assumptions, if the physician is predisposed to prescribing the generic drug, then it is not possible for the pharmaceutical company to completely dissuade the physician from doing so. However, as we demonstrate in our example of Section 2, if the physician is ambiguity averse, in particular, if she is maxmin expected-utility maximizers à la (Gilboa and Schmeidler, 1989), and if the pharmaceutical company has ambiguous tests at his disposal, he is able to achieve just that.

For clarity, we refer to Kamenica and Gentzkow's (2011) communication devices/signals as *probabilistic* devices. In our model, we introduce *ambiguous* devices. An ambiguous communication device is a set of probabilistic devices. Upon reception of a message, Receiver updates her (unique) prior with the Full Bayesian rule (à la Pires 2002 and Epstein and Schneider 2007), i.e., she updates her prior with respect to each *probabilistic* device, which yields a set of posterior beliefs for each message. Sender and Receiver are ambiguity averse à la Gilboa and Schmeidler (1989)–when evaluating an action/ambiguous device, they compute their expected utility for any possible posterior belief and rank actions/communication devices according to their minimum expected utility.

The use of ambiguous devices carries both opportunities and challenges. First, Sender benefits from increased leeway regarding how to control the information flow to Receiver. Therefore, he can induce Receiver to act in such a manner that would not be feasible with probabilistic devices alone. On the other hand, Sender introduces ambiguity where there was initially none, which would generically decrease his ex ante utility (given that he commits to a device before learning anything). It is therefore unclear a priori whether the expert can strictly benefit from ambiguity.

In this paper, we provide a characterization of Sender's optimal payoff under ambiguous persuasion. First, we present a "splitting lemma," namely, the possible profile of posterior sets that are achievable with ambiguous signals, which precisely pins down the extent to which ambiguous signals expand beyond the Bayesian plausibility condition given by Kamenica and Gentzkow (2011). Second, we demonstrate that, if Sender uses ambiguous signals in persuasion, every probabilistic device he uses under optimal persuasion must give him the same expected payoff. He achieves this through synonyms, which are messages that induce the same set of posteriors and therefore the same action from Receiver. If two probabilistic devices did not generate the same payoff, then he could use synonyms to approximate the higher payoff of the two, and therefore improve his expected payoff. In effect, Sender creates a set of synonyms to hedge himself against his own ambiguity. Finally, we show that Sender's optimal payoff can be characterized by the concave closure of a value function over the set of posteriors, which is defined as the *highest possible* expected payoff Sender may achieve according to the posterior and given that the posterior belongs to Receiver's posterior set. Our result is reminiscent of the "concavification" characterization of Kamenica and Gentzkow (2011), but the precise upper bound of the concave closure we find is somewhat surprising, which brings insights that are quite apart from Kamenica and Gentzkow's.

We then proceed to explore the structure of optimal ambiguous signal in two special applications.

In the first application, we consider the case when Sender has a most preferred action, which is also "safe" for Receiver, in that Receiver could always do worse than take this action in some state of the world. Even though this action might be ex ante worse than another action, through ambiguous messages, Sender can succeed in persuading Receiver to always take this safe action, which is not possible under standard Bayesian persuasion. The reason is that, with ambiguous signals, Sender can convince Receiver that she should be worried about a bad state occurring if she chose a different action than the safe one.

In the second application, we consider the frequently studied "uniform-quadratic" case made popular by Crawford and Sobel (1982). We characterize the optimal simple ambiguous signal structure, where each signal realization is associated with two possible posteriors, both of which are uniform distributions over intervals. The optimal simple ambiguous signal structure features an equal partition of state space into finitely many unambiguous intervals and maximal ambiguity within each unambiguous interval. This is very different from the conclusion from standard Bayesian persuasion, where Sender finds it optimal to perfectly reveal all the information to Receiver, despite the conflict of interest between them. This provides a justification for ambiguity of communication even when Sender appears to have commitment power (for example, rating agencies, who are long-run players and have relatively stable rating categories).

Related Literature

In this paper, we adopt the full-commitment assumptions of Kamenica and Gentzkow (2011) but extend their model to study ambiguous communication devices. In this respect, two recent papers by Laclau and Renou (2016) and Lipnowski and Mathevet (2015) are relatively closely related to our work. Laclau and Renou (2016) consider *public persuasion*, where Sender sends the same signal to multiple receivers with heterogeneous prior beliefs, which can be alternatively interpreted as a Bayesian sender persuading an ambiguity-averse receiver with multiple priors.[4] They characterize a splitting lemma, the counterpart of Kamenica and

[4] Their model can be viewed as an extension of Alonso and Câmara (2016), who introduce into Kamenica and Gentzkow's (2011) model disagreement over priors between Sender and Receiver. Alonso and Câmara (2016) identify conditions under which Sender benefits from persuading Receiver.

Gentzkow's (2011) concept of Bayes plausibility, and provide a version of "concavification" in their environment. We differ from their approach by focusing on the case when there is no prior ambiguity while the information technology is ambiguous.We also assume both Sender and Receiver can be ambiguity-averse to ensure symmetric information. Lipnowski and Mathevet (2015) consider a model of Bayesian persuasion where Receiver has psychological preferences, where beliefs directly enter Receiver's payoff function. They demonstrate that Sender might not want to fully reveal her information to Receiver even if there is no intrinsic conflict of interest between them. They demonstrate that the revelation principle might fail to hold (in Section 5 of their paper), when Receiver's preferences depend on beliefs in a way that violates expected utility maximization.

We demonstrate in our applications that Sender may want to commit to ambiguous signals, even if Sender is himself ambiguity averse and even if he would like to fully reveal information to Receiver absent ambiguous signals. Thus, our results can be viewed as to provide a new justification for the widespread use of vague language in interpersonal and organizational communication. Previous work has focused on cheap-talk communication in the manner of Crawford and Sobel (1982), where Sender does not have commitment power (see Sobel (2013) for a review). Blume et al. (2007) and Blume and Board (2014) show that the presence of vagueness may facilitate communication between Sender and Receiver. Kellner and Le Quement (2015) solve a simplified two actions/two states game where ambiguity is present in the receiver's priors but not as a strategic choice of the sender. They show that the sender would use more messages under this assumption than with the regular Bayesian prior. In a follow-up work, Kellner and Le Quement (2016) introduce endogenous ambiguous messages into the cheap-talk framework of Crawford and Sobel (1982). They demonstrate that the possibility of ambiguous messages, coupled with ambiguity aversion of Receiver, may improve the communication between Sender and Receiver. These two papers' approach differ from ours in that they do not assume that the sender can commit. Lipman (2009) argues that it is puzzling that vagueness of language pervades in seemingly common-interest situations. To a certain extent, our results in the uniform-quadratic example demonstrate that a common-interest situation with expected utility maximizers, in the sense that Sender and Receiver both prefer that information be fully revealed *under commitment*, can easily turn into one that is not.[5]

Recent experimental research has supported the hypothesis that people are ambiguity averse—they dislike betting on an event with unknown probability.[6] This phenomenon has attracted

[5] Crémer et al. (2007) explain the use of vague language as a consequence of the cost of being precise. See also Sobel (2015). Ivanov (2010) shows that if the decision maker can control what information an expert can get before communication, then she would give the expert relatively coarse information that also takes an interval structure.

[6] For a review of earlier experimental evidence, see Camerer and Weber (1992). For more recent experiments, see for instance Fox and Tversky (1995), Chow and Sarin (2001), Halevy (2007), Bossaerts et al. (2010), and Abdellaoui et al. (2011).

significant interest in theory and applications.[7] There has been a growing theoretical literature exploring the role of ambiguity aversion in games and mechanisms (Bose et al., 2006; Lopomo et al., 2011; Wolitzky, 2016; Lopomo et al., 2014; Frankel, 2014; Di Tillio et al., 2015; Ayouni and Koessler, 2015). Bade (2011) and Riedel and Sass (2014) both study complete information games where the players, in addition to playing mixed strategies, are allowed to use ambiguous strategies in a manner similar to ours. Most relevant to ours, Bose and Renou (2014) introduce similar ambiguous devices in a communication stage preceding a mechanism design problem. The key insight is the designer can cleverly use these devices to exploit the ambiguity aversion of the agents. That paper shows that a wider set of social choice functions are implementable with ambiguous communication devices.

The rest of our paper is structured as follows: Section 2 illustrates how ambiguous communication devices can be used to benefit the sender. Section 3 introduces the framework and presents the communication game. Section 4 characterizes the optimal value of ambiguous persuasion as the concave closure of the sender's conditional utility function. Subsection 4.1 provides a necessary and sufficient condition for the set of posterior beliefs to be plausibly updated via some ambiguous device. Subsection 4.2 introduces a novel concept called synonyms shows how an expert can make use of them to hedge himself against ambiguity. Section 5 provides two examples in which ambiguous communication device with synonyms improve upon Bayesian persuasion. Section 6 discusses some additional constraints on ambiguous devices such as dynamic consistency and positive value of information from ambiguous communication. Section 7 concludes. Proofs omitted in the main text are relegated to the Appendix.

2 An Illustrating Example

We first present an example that demonstrates how ambiguous persuasion improves upon Bayesian persuasion. Let Sender be a brand name drug producer and Receiver be a physician. The physician could choose between two actions: prescribe the brand name producer's drug or prescribe a generic competitor of the drug. The brand name drug producer always prefers that the physician prescribe the brand name drug, but the physician's preference depends on how effective the generic drug is. Suppose (according to common belief) the brand name drug is less risky than the generic drug. From the physician's point of view, we assume for simplicity that the brand name drug is always effective and produces some deterministic treatment utility u_H; the generic drug's effectiveness is uncertain: in one state (denoted as $\omega =$ Effective), it is as effective as brand name drug and produces the same high utility u_H; in the other state (denoted as $\omega =$ Ineffective), it is low quality and pro-

[7] Gilboa and Marinacci (2013) survey the vast literature of axiomatic foundations for ambiguity aversion. Mukerji and Tallon (2004) and Epstein and Schneider (2010) survey economic and financial applications of ambiguity aversion.

duce low utility u_L ($< u_H$) (due to unintended side effect, less effective).[8] Normalize and assume the unit price of the generic drug is 0 and that of the brand name drug is $c > 0$ (utils). Assume the physician is well-intended and so her payoff is treatment utility from the drug minus its cost; while Sender, the brand name drug producer, always prefers its own drug prescribed. The following matrix summarizes the state-and-action-contingent payoffs to Sender and Receiver. In each entry, the first element is Sender's payoffs and the second element is Receiver's payoffs.

	Effective	Ineffective
Brand name	$(1,\ u_H - c\ )$	$(1,\ u_H - c\ )$
Generic	$(0,\ u_H\ )$	$(0,\ u_L)$

The physician and the brand name drug producer share some common prior $p(\text{Effective}) = p_0 \in (0, 1)$. Suppose $c/(1 - p_0) > u_H - u_L > c > 0$, and so the physician strictly prefers prescribing the generic drug without additional information.

First, suppose the brand name producer can only commit to Bayesian messages that are on average correct. Then by Kamenica and Gentzkow's (2011) Proposition 4 and 5, the optimal Bayesian communication device is a tuple (M, π^*) with the following properties: (i) It suffices to have two messages—$M = \{e, i\}$, where $m = e$ corresponds to the message "generic is effective" and $m = i$ "generic is ineffective." (ii) The mapping from states to messages is

$$\begin{aligned} \pi(e|\text{ Effective}) &= \frac{p - p^*}{p(1 - p^*)}, \qquad \pi(i|\text{ Effective}) = \frac{(1 - p)p^*}{p(1 - p^*)}; \\ \pi(e|\text{ Ineffective}) &= 0, \qquad \pi(i|\text{ Inffective}) = 1, \end{aligned}$$

which induces a Bayes plausible distribution over posteriors

$$\begin{aligned} p_e(\text{Effective}) &= 1, \\ p_i(\text{Effective}) &= p^*, \end{aligned}$$

where p^* is the posterior that makes her exactly indifferent between the brand name and the generic drugs,[9] that is,

$$p^* = 1 - \frac{c}{u_H - u_L}.$$

Note that

$$Prob(m = i) = \frac{1 - p}{1 - p^*} < 1.$$

[8]See, for example, a 2013 New York Times article covering the Ranbaxy (an Indian drug company that had sold generic versions of brand name drugs, say Lipitor, in the US market) fraud scandal, which raises quality concerns about generic drugs (http://www.nytimes.com/2013/05/14/business/global/ranbaxy-in-500-million-settlement-of-generic-drug-case.html).

[9]By assumption, $p > p^*$.

Under Bayesian persuasion, the physician will prescribe the brand name drug if and only if the message is in favor of the brand name ($m = i$), which occurs with probability $Prob(m = i) \in (0, 1)$.

In contrast, when ambiguous messages are allowed, the brand name producer can persuade the physician (to prescribe the brand name drug) with probability one. The idea is to communicate in a way such that the messages have multiple ways of interpretations, with the requirement that each way of interpretations is correct on average. For example, the brand name producer could sponsor an experimental test of the generic drug and report the data, while remaining vague on how data are generated and therefore how they should be interpreted. A simple formalization of this idea is a system with $M = \{e, i\}$ that admits two likelihood distributions,[10] $\Pi = \{\pi, \pi'\}$, where

$$\pi(m = i|\text{ Inffective}) = 1, \qquad \pi(m = e|\text{ Effective}) = 1;$$
$$\pi'(m = i|\text{ Effective}) = 1, \qquad \pi'(m = e|\text{ Inffective}) = 1.$$

Following Epstein and Schneider (2007), we assume the physician is a maxmin EU (MEU) maximizer and updates her beliefs likelihood-by-likelihood for all elements in Π. Hence for all prior $p_0 \in \{0, 1\}$, she forms sets of posterior:

$$\begin{aligned} P_i(\text{Effective}) &= \{0, 1\}, \\ P_e(\text{Effective}) &= \{0, 1\}. \end{aligned}$$

As we will verify later, these posteriors are also "Bayes plausible." Therefore, the MEU physician will prescribe the brand name drug no matter what signal she observes. In this way, the brand name producer does strictly better by using the ambiguous communication device than using the optimal Bayesian one. Moreover, this simple communication device with maximal ambiguity is optimal, as it induces the Sender optimal action with probability one.

3 Framework

We consider a persuasion game between a sender and a receiver. Let Ω denote the set of states of the world and A be the set of actions. Assume Ω and A are compact subsets of Euclidean space. Receiver and Sender have continuous utility functions, denoted by $u(a, \omega)$ and $v(a, \omega)$, respectively. Sender and receiver share a common prior $p_0 \in \Delta(\Omega)$ with full support.[11]

[10] We use a likelihood distribution to model an interpretation of the signal system.

[11] Let X be a compact subset of a metric space, endowed with the Borel topology. Throughout $C(X)$ to denote the set of real-valued functions on X, and $\Delta(X)$ to denote the set of probabilities on X endowed with the topology of weak convergence. The space of closed and convex subsets of $\Delta(X)$ is endowed with the standard Hausdorff topology. Let Y be a finite subset of a metric space, we use $co(Y)$ to denote the convex hull of Y.

Sender sends some message in the finite set M to Receiver, who then takes an action. A *probabilistic* (or Bayesian) communication device π is a function from states of the world to lotteries over messages, where $\pi(\cdot|\omega) \in \Delta M$. We denote $\pi(m|\omega)$ the probability that message m is sent at state ω and $\pi(m) = \sum_{\omega} p_0(\omega)\pi(m|\omega)$ the probability that message m is sent. An *ambiguous* communication device Π is the convex hull of a finite family of probabilistic devices that all use the same messages. That is, there exists a finite family of probabilistic devices $(\pi_k)_K$ indexed by $k \in K = \{1, .., |K|\}$ such that $\Pi = co((\pi_k)_K) = \left\{\pi \in \Delta(M)^{\Omega} \;\middle|\; \exists\lambda \in \Delta(K), \pi = \sum_{k\in K} \lambda(k)\pi_k\right\}$ and for all k and $j \in K$, $\pi_k(m) > 0 \Rightarrow \pi_j(m) > 0$. Without loss, assume $(\pi_k)_K$ are the K distinct extreme points of set Π.[12]

Following Kamenica and Gentzkow (2011), the persuasion game has two stages. Ex ante, Sender commits to an ambiguous communication device (M, Π). One interpretation is that Sender designs an Ellsberg urn with K numbered balls and delegates to a third party, who then chooses privately a ball from this urn. If the k-th ball is chosen, then the (M, π_k) communication device will generate the message. The players are ignorant about how the third party draws the ball. In the second (ex post) stage, both players observe the realized message and Receiver takes an action a. Then the true state is revealed and payoffs are realized. The game is solved by backward induction with a sender-preferred tie-breaking rule: (i) In the second stage, Receiver forms a set of posteriors P_m and takes an action to maximize her ex post maxmin expected utility; If there is a tie, Receiver chooses the most preferred act by Sender. (ii) In the first stage, Sender chooses possibly ambiguous message structure (M, Π) that to maximizes his ex ante maxmin expected utility.

3.1 Beliefs and Preferences

Fix prior p_0 and some ambiguous communication device (M, Π) with multiple likelihoods. We specify the players' ex ante and ex post beliefs and preferences.

In the ex post stage, we follow Epstein and Schneider (2007) and assume Sender and Receiver update beliefs likelihood by likelihood.[13] That is, upon receiving message m, Sender and Receiver's set of posteriors is

$$co(P_m) = co\{p_m^{\pi_k} \in \Delta(\Omega) : p_m^{\pi_k}(\omega) = \frac{p_0(\omega)\pi_k(m|\omega)}{\int_{\omega'} \pi_k(m'|\omega')p_0(\omega')d\omega'}, \quad \forall \pi_k \in (\pi_k)_K, \omega \in \Omega\},$$

where $p_m^{\pi_k}$ is the m-Bayesian-posterior induced by probabilisitic device π_k.

[12]Throughout the paper, we do not distinguish between the set of K probabilistic devices and its convex hull, because only the extreme points of the set of priors matter for a maxmin EU agent.

[13]See Epstein and Seo (2010) for an axiomatization of this likelihood-by-likelihood updating rule.

With ambiguous communication, the set of posteriors may become less precise than the prior *at all realized messages.* As illustrated by the following example (Seidenfeld and Wasserman, 1993), this is a natural consequence of updating with multiple probabilities.

Example 1 (Dilation)**.** Consider a fair coin that will be tossed twice. Let H_i and T_i denote the event that the i-th coin toss lands Head and Tail, respectively ($i = 1, 2$). The unconditional probability of the second coin toss landing HEAD is clearly $\frac{1}{2}$, that is, $p_0(H_2) = \frac{1}{2}$. However, the players don't know what techniques are used in between the two tosses and how they are related (for e.g., the coin might or might not be flipped to the other side after the first toss). Hence the players' belief about the joint probability of two HEADs, $\Pr(H_1 \text{ and } H_2)$, can lie arbitrarily in the interval $[0, \frac{1}{2}]$. Suppose now the players observe the outcome of the first coin toss and are asked about their posterior beliefs about the event H_2. Then the posteriors are $P_{H_1}(H_2) = P_{T_1}(H_2) = [0, 1]$. Note the players' posterior beliefs about the event H_2 are dilated away from the prior for either realization of the first coin toss.

Receiver's decision criterion is her ex post maxmin EU. Upon observing m, Receiver's utility is

$$U(a, P_m) = \min_{p_m \in co(P_m)} E_{p_m}[u(a, \omega)] = \min_{p_m \in P_m} E_{p_m}[u(a, \omega)]^{14}$$

Her optimal action is

$$a^*(m) \in \arg\max_{a \in A} \min_{p_m \in P_m} E_{p_m}[u(a, \omega)].$$

Let $\hat{a}(P_m)$ denote the optimal action taken by Receiver under posterior beliefs P_m and the Sender-preferred tie-breaking rule.

In the first stage, Sender chooses an ambiguous communication device (M, Π) to maximize his *ex ante* maxmin EU. Sender's value in the persuasion game at prior p_0 is [15]

$$\sup_{(M,\Pi)} v(p_0, \Pi) = \sup_{(M,\Pi)} \min_{\pi \in \Pi} E_{p_0}[E_\pi[v(\hat{a}(P_m), \omega)|\omega]]. \tag{1}$$

[16] We use V_0 to denote Sender's utility without communication: $V_0 = E_{p_0} v(\hat{a}(p_0), \omega)$.

[14]Since we focus on maxmin EU, a linear minimization problem subject to a convex set of beliefs, it is without loss of generality to focus on the extreme points of the belief set. The same argument applies to Sender's Maxmin EU.

[15] Note that with ambiguous beliefs and our assumption of updating maxmin EU preferences, Sender's ex ante and ex post preferences might not be dynamically consistent. We follow the solution concept proposed by Siniscalchi (2011), Stroz-type *consistent planning*, to address dynamic inconsistencies in a decisionmaker's preferences—the decisionmaker considers all contingent plans that will actually be carried out by his future preferences and chooses among them the one that is optimal according to his current preferences. Applied to our setting, consistent planning requires Sender to consider all actions that will be taken by Receiver ex post and, going backward, choose a signal structure according to Sender's ex ante MEU preferences.

[16]Precisely, p_0 and Π induce a set of joint probabilities over the product set $\Omega \times M$. Let it be $P_{0,\Pi} := co\{p \in \Delta(M \times \Omega) : p(m, \omega) = p_0(\omega)\pi(m|\omega) \quad \forall \pi \in \Pi\}$. Sender has maxmin EU with prior set $co(P_{0,\Pi})$ and utility index $v(\hat{a}(P_m), \omega)$.

We can also express an ambiguous communication device Π as an induced set of distributions over posteriors, generalizing the technique by Kamenica and Gentzkow (2011). Let $\mathbf{p} = (p_m)_{m\in M}$ be a profile of posteriors at each signal and $\tau \in \Delta(M)$ be a marginal distribution over M. Then the set of distributions over posteriors generated by the ambiguous device Π is

$$R = co\{(\tau_k, \mathbf{p}_k) \in \Delta(M)\times(\Delta\Omega)^M : \tau_k(\cdot) = \int_\Omega p_0(\omega')\pi_k(\cdot|\omega')d\omega', \mathbf{p}_k = (p_m^{\pi_k})_{m\in M}, \forall \pi_k \in (\pi_k)_K\}.$$

Clearly, each $(\tau_k, \mathbf{p}_k) \in R$ is Bayes plausible so that $\sum_{m\in M} \tau_k(m)p_{mk} = p_0$. Moreover, for any convex set with finitely many extreme points $R \subseteq \Delta(M) \times (\Delta\Omega)^M$ satisfies *generalized Bayes plausibility* if every $(\tau, \mathbf{p}) \in R$ is Bayes plausible.

We can rewrite Sender's ex-ante utility as

$$v(p_0, \Pi) = \min_{(m_k, \mathbf{p}_k)\in R} E_m[E_{p_m}[v(\hat{a}(P_m), \omega)|m]]. \tag{2}$$

The function $v_{p_m}(P_m) := E_{p_m}[v(\hat{a}(P_m), \omega)|m]$ is the Sender's expected utility when computed with respect to some Sender's posterior $p_m \in P_m$ and Receiver's best response $\hat{a}(P_m)$, which will be very useful for us.

4 A characterization of the value of persuasion

The introduction of ambiguous devices allows Sender new ways to influence Receiver's action. It is straightforward to see that Sender must be at least as well-off as he is without ambiguous devices. One the one hand, the opportunity to Sender appears limitless. Yet, there must exist *some* limit to what Sender can achieve. On the other hand, it is not clear whether ambiguous persuasion can *strictly* benefit Sender. After all, Sender himself is ambiguity averse and may suffer from any endogenous ambiguity he introduces into persuasion. In the illustrating example in Section 1, we have demonstrated that it it indeed possible for Sender to strictly benefit from ambiguous persuasion. In that example, Sender is able to achieve the highest payoff possible from his interaction with Receiver. But, a simple example below can demonstrate that this is not always the case. And our main result (Proposition 1) provides a characterization of the optimal value Sender can possibly achieve by using ambiguous devices.

In Kamenica and Gentzkow (2011), the value of optimal Bayesian persuasion is given by $\hat{V}(p_0)$ where $\hat{V}$ is the concave closure of the function defined by $v(p) = \mathbb{E}_p v(a^*(p), \omega)$ the expected payoff of Sender when Receiver holds the Bayesian posterior belief $p \in \Delta(p_0)$. In this section, we define an equivalent function for the case of ambiguous persuasion.

Let K be a finite number of Bayesian devices to be used. For any given vector of distributions $P = (p_k)_{k\in K}$, let $v_k(P) = \mathbb{E}_{p_k} v(a^*(P), \omega)$ be the payoff of Sender when Receiver holds ambiguous posterior beliefs P and when computed with respect to the k-th distribution in the posterior vector.[17]

Consider the concave closure of v_k:

$$V_k(P) = \sup\{z | (P, z) \in co(\text{Graph}(v_k))\}.$$

where $co(\text{Graph}(v_k))$ is the convex hull of the graph of function v_k, that is, $co(\{(P, z') \in (\Delta\Omega)^{MK} \times \mathbb{R} | v_k(P) \geq z'\})$. By definition, if $(P, z) \in co(\{(P, z') | v_k(P) \leq (\geq ?) z'\})$, then there exists a distribution of posteriors $\tau \in \Delta M$ and a vector of posterior sets $\{P_m\}_{m \in M} \in (\Delta\Omega)^{MK}$ such that $\mathbb{E}_\tau P_m = P$ and $\mathbb{E}_\tau v_k(P_m) = (\geq ?) z$.

Given $V_k = V_j$ (for $k, j \in K$) by construction, we will now focus only on V_1 without loss of generality. Consider now the maximal projection of $V_1(P)$ over $\Delta\Omega$, that is, let

$$\overline{V}(p) = \max_{P^{-1} \in (\Delta\Omega)^{K-1}} V(co(p, P^{-1})).$$

Our main result is that the latter function characterizes the optimal value of ambiguous persuasion. We consider that Sender benefits from ambiguous persuasion, as opposed to Bayesian persuasion, if there exists an ambiguous device such that its value is greater than any Bayesian device.

Proposition 1. *Sender benefits from ambiguous persuasion if and only if* $\overline{V}(p_0) > \hat{V}(p_0)$.

We show below an example where we compute the $\overline{V}$ function and derive the value of ambiguous persuasion according to proposition 1.

Example 2. Let there be two states, low and high, denoted by w_l and w_h, and three actions, low, middle, and high, denoted by a_l, a_m, and a_h, respectively. In the payoff matrix in Table 1, the first number of each cell is Sender's payoff and the second that of Receiver. Given equiprobable low and high states in prior belief, the default action of Receiver is the middle and safe action, which yields a payoff of 0 to Sender.

We first look at how Sender benefits from Bayesian persuasion. Figure 1 gives the concave closure of Sender's payoffs in the Bayesian case. This provides the value of Bayesian persuasion as well as the posteriors needed to attain this value. As seen in Figure 1, the

[17] For simplicity of notation, we make no distinction between the vector of posteriors $P = (p_k)_{k \in K}$ and the set of posteriors that is obtained from collapsing the vector by removing posteriors that are redundant, noting that Receiver's payoff depends only on the latter.

	ω_l	ω_h
a_l	$(-1, 3)$	$(-1, -1)$
a_m	$(0, 2)$	$(0, 2)$
a_h	$(1, -1)$	$(1, 3)$

Table 1: Payoff matrix for Example 2.

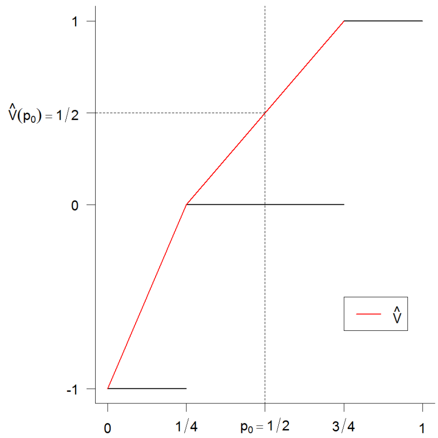

Figure 1: The value of probabilistic communication. The horizontal axis is the posterior probability of the low state and the vertical axis is Sender's payff.

optimal Bayesian device should yield posteriors of 1/4 and 3/4 for a value $\hat{V}(p_0) = 1/2$. Let the first message be called "high" and second one "middle," each sent with probability 1/2. When receiving the message "high," Receiver is indifferent between a_h and a_m and takes the former (Sender-preferred) action; when receiving the message "middle," Receiver is indifferent between a_l and a_m and takes the latter (Sender-preferred) action.

With two Bayesian devices used, Receiver's best response function $a^*(p_1, p_2)$ can be written

$$a^*(p_1, p_2) = \begin{cases} a_h, & \min\{p_1, p_2\} \geq 3/4; \\ a_l, & \max\{p_1, p_2\} < 1/4; \\ a_m, & \text{otherwise.} \end{cases}$$

Therefore, Sender's expected payoff based on his first posterior can be written

$$v_1(p_1, p_2) = \begin{cases} -1, & \min\{p_1, p_2\} \geq 3/4; \\ 0, & \max\{p_1, p_2\} < 1/4; \\ 1, & \text{otherwise.} \end{cases}$$

We now turn to computing the value of ambiguous persuasion to Sender when two Bayesian

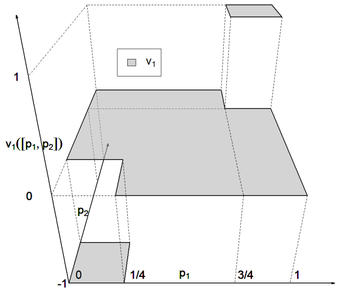

Figure 2: $v_1(P)$

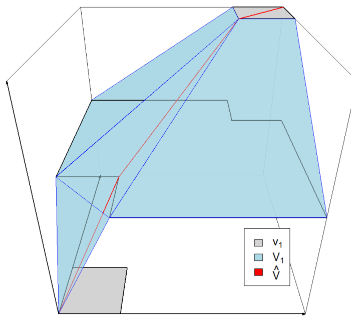

Figure 3: $V_1(P)$

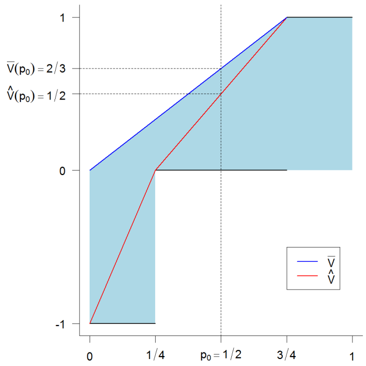

Figure 4: $\overline{V}(p)$

devices are used, which as we show in Proposition 1, is characterized by the function $\overline{V}$. Figures 2, 3, and 4 illustrate the step-by-step construction of the $\overline{V}$ function.

From these figures, we have that $\overline{V}(p_0) = 2/3$. From proposition 1, Sender can therefore benefit from ambiguous persuasion.

The full proof of this result is given in appendix A.4. This proof makes use of two intermediate results which we will introduce in the next two subsections.

4.1 Splitting lemma

In this subsection, we show that finding an optimal ambiguous device is equivalent to finding an appropriate profile of posterior beliefs.

For fixed prior p_0 and message M, we want to know if some profile of (closed and convex) posterior sets $\{P_m\}_{m \in M} \subseteq \Delta(\Omega)^M$ can be induced by a prior p_0 and some ambiguous communication device Π.

We say a profile of posterior sets $(P_m)_{m \in M} \subseteq (\Delta(\Omega))^M$ is *potentially generalized Bayes plausible* if for arbitrary $m \in M$ and $p_m \in P_m$, there exists a profile of posteriors $(p'_{m'})_{m' \in M} \in (P_{m'})_{m' \in M}$ such that $p'_m = p_m$ and $\sum_{m'} \tau_{m'} p'_{m'} = p_0$ for some nonnegative weights $0 \leq \tau_{m'} \leq 1$, $\sum_{m'} \tau_{m'} = 1$.

A profile of posterior sets $(P_m)_{m \in M}$ *is likelihood-by-likelihood updated by some ambiguous communication device*, if there exists some device Π such that for all $m \in M$, $P_m =$

$\cup_{\pi\in\Pi}\{p^{\pi}_{0,m}\}$ with $p^{\pi}_{0,m}(\omega) = \frac{\pi(m|\omega)p_0(\omega)}{\int_{\Omega}\pi(m|\omega')p_0(\omega')d\omega'}$ being the posterior of the Baysian device π.

The next lemma shows the equivalence of these two concepts. This could be viewed as the multiple likelihood analogy of the Bayes plausibility condition (Kamenica and Gentzkow, 2011) or the splitting lemma (Aumann and Maschler, 1995).

Lemma 1. *Fix prior p_0 and message set M. There exists some ambiguous communication device such that $(P_m)_{m\in M}$ is the likelihood-by-likelihood updated profile of posterior sets if and only if $(P_m)_{m\in M}$ is potentially generalized Bayes plausible.*

Lemma 1 says that not every selection of posterior profile could correspond to the updated posteriors by some ambiguous communication device. It puts a bound on the set of generalized Bayes plausible posteriors that Sender could credibly induce. Clearly Sender has more leeway to induce the desirable posteriors than what is constrained by Bayes plausiblility, but this freedom is not without limit. This is illustrated by the following example.

Example 3. Consider for instance Example 2. As in (Kamenica and Gentzkow, 2011), our construction is useful to determine sets of posteriors that we would want to generate from an ambiguous device. In this case, two messages are needed. We denote them m_l and m_h. The first device would need to yield posteriors of $p_{1,m_l} = 0$ and $p_{1,m_h} = 3/4$. To define posteriors for a second Bayesian device, one needs to determine the posterior beliefs needed to have ensured the decision maker took the correct actions at messages 1 and 2, that is, distributions in $\arg\max(v_k([p_1, p]))$ and $\arg\max(v_k([p_1, p]))$.

In this example, we will take $p_{2,m_l} = 1/4$ and $p_{2,m_h} = 3/4$. This therefore provides two sets of posteriors:

$p_1(\omega\|m)$	ω_l	ω_h
m_l	1	0
m_h	$\frac{1}{4}$	$\frac{3}{4}$

$p_2(\omega\|m)$	ω_l	ω_h
m_l	$\frac{3}{4}$	$\frac{1}{4}$
m_h	$\frac{1}{4}$	$\frac{3}{4}$

Those two sets are potentially generalized Bayes plausible as $p_0 = 1/2 \in [0, 3/4]$ and $p_0 \in [1/4, 3/4]$. And indeed, it is possible to find two Bayesian devices which would lead to such posteriors:

$\pi_1(m\|\omega)$	ω_l	ω_h
m_l	$\frac{2}{3}$	0
m_h	$\frac{1}{3}$	1

$\pi_2(m\|\omega)$	ω_l	ω_h
m_l	$\frac{3}{4}$	$\frac{1}{4}$
m_h	$\frac{1}{4}$	$\frac{3}{4}$

Furthermore, note that if Sender were to use the ambiguous device Π derived from those two, then receiver will choose a_m upon seeing m_l and a_h upon seeing m_h. And $\tau_1 = (2/3, 1/3)$.

So we have that $v_1(p_1, p_2) = 2/3 = \overline{V}(p_0)$.

Note however that our construction does not always lead naturally to potentially generalized Bayes plausible posterior sets. Consider for instance the same payoffs but assume the prior is now equal to 1/8. The interesting set of posteriors we would like to generate are still $[0, 1/4]$ and $\{3/4\}$. However, those sets are not potentially generalized Bayes plausible given the prior of 1/8. Indeed, $1/8 \notin [1/4, 3/4]$. Lemma 1 therefore suggests that no device could lead to such sets of posteriors.

We can however consider completing our posterior sets with a third message in order to ensure potentially generalized Bayes plausibility. First, note that a profile of posterior distributions will by construction be Bayes plausible (those resulting from the concave closure). Only those posterior profile resulting from the projection might not be Bayes plausible. As a result, we propose to complete our posterior sets by $\{p_0, p\}$ where p is chosen so that $co(p, 1/4, 3/4)$ would be Bayes plausible.

The posteriors generated by the two devices, listed below, are potentially generalized Bayes plausible.

$p_1(\omega\|m)$	ω_l	ω_h
m_l	1	0
m_h	$\frac{1}{4}$	$\frac{3}{4}$
m_r	$\frac{7}{8}$	$\frac{1}{8}$

$p_2(\omega\|m)$	ω_l	ω_h
m_l	$\frac{3}{4}$	$\frac{1}{4}$
m_h	$\frac{1}{4}$	$\frac{3}{4}$
m_r	1	0

There exists therefore a pair of Bayesian devices that can lead to these sets of posteriors from $p_0 = 1/8$. Furthermore, by adding p_0 to the additional posterior set, we ensure that the first Bayesian device considered can use the same relative weights on m_l and m_h as without this third posterior set. The following two devices lead to the three posterior sets defined here.

$\pi_1(m\|\omega)$	ω_l	ω_h
m_l	$\frac{20\varepsilon}{21}$	0
m_h	$\frac{\varepsilon}{21}$	ε
m_r	$1-\varepsilon$	$1-\varepsilon$

$\pi_2(m\|\omega)$	ω_l	ω_h
m_l	$\frac{3}{28}$	$\frac{1}{4}$
m_h	$\frac{1}{28}$	$\frac{3}{4}$
m_r	$\frac{6}{7}$	0

Note that if Sender were to use the ambiguous device Π derived from those two Bayesian ones, then Receiver will choose a_m upon seeing m_l, a_h upon seeing m_h, and a_l upon seeing m_r.

For the second device, a Bayes plausible distribution over posterior is $\tau_2 = (1/8, 1/8, 6/8)$. In this case, $v_2(p_1, p_2) = 1/8 * 1 - 6/8 * 1 = -5/8$.

For the first device, we have that $\lim_{\varepsilon \to 1} \tau^1 = (5/6, 1/6, 0)$ and $\lim_{\varepsilon \to 1} v_1(p_1, p_2) = 5/6*(0)+1/6*1 = 1/6 = \overline{V}(1/8)$.

Coupled with Lemma 1, this example shows how one can construct an ambiguous device whose value to Sender is *at least* arbitrarily as close to $\overline{V}(p_0)$. The next subsection shows how Sender can use *synonyms*, messages that lead to the same posterior sets, in order to hedge against ambiguity and achieve the value as computed by the best Bayesian device.

4.2 The use of synonyms

In this section, we show that Sender can hedge himself against ambiguity by using what we call synonyms, that is several messages that yield the same posterior beliefs for Receiver.

First, I define the $\oplus$ operation over probabilistic devices. Consider two devices π_1 and π_2 which use the same messages in M_1. Let M_2 be a duplicated set of messages: $M_1 \cap M_2 = \emptyset$ and there exists a bijection b between M_1 and M_2. Given $\alpha \in [0, 1]$, let $\pi' = \alpha\pi_1 \oplus (1-\alpha)\pi_2$ be the device that sends a message $m_1 \in M_1$ with probability $\alpha\pi_1(m_1/\omega)$ and a message $m_2 = b(m_1) \in M_2$ with probability $(1-\alpha)\pi_2(m_1/\omega)$ from state ω. One interesting feature of the $\oplus$ operation is that the value of probabilistic devices is linear with regards to it.

Lemma 2. *The value function is linear with respect to the $\oplus$ operation:* $V(\alpha\pi_1 \oplus (1-\alpha)\pi_2) = \alpha V(\pi_1) + (1-\alpha)V(\pi_2)$.

Proof in Subsection A.2.

Consider now an ambiguous device $\Pi = conv((\pi_1, \pi_2))$. Define $\Pi' = co((\pi_1', \pi_2'))$ the ambiguous device such that $\pi_1' = \alpha\pi_1 \oplus (1-\alpha)\pi_2$ and $\pi_2' = (1-\alpha)\pi_2 \oplus \alpha\pi_1$. This yields the following ambiguous device:

Π'	π_1'	π_2'
$m_1 \in M_1$	$\alpha\pi_1(m_1/\omega)$	$(1-\alpha)\pi_2(m_1/\omega)$
$b(m_1) \in M_2$	$(1-\alpha)\pi_2(m_1/\omega)$	$\alpha\pi_1(m_1/\omega)$

The posterior set of beliefs induced by Π' are the same for messages m_1 and $b(m_1)$. In that sense, m_1 and its equivalent $m_2 = b(m_1)$ are *synonyms*. Furthermore, these posterior sets are the same as those induced by the original ambiguous device Π. As a result, Π' and Π induce the same actions from Receiver.

Finally, using Lemma 2, the value of Π' is $V(\Pi') = \alpha V(\pi_1) + (1-\alpha)V(\pi_2)$. Assuming for example that $V(\pi_1) > V(\pi_2)$, then by picking α ever closer to 1, the value of Π' converges to that of π_1. From this follows the subsequent Corollary:

Corollary 1. *Given an ambiguous device* $\Pi = co((\pi_k)_K$*, there exists a sequence of devices* $(\Pi'_{k'})_{k' \in \mathbb{N}}$ *using synonyms such that* $\underset{k' \to +\infty}{lim} V(\Pi'_{k'}) = \underset{\pi \in \Pi}{sup} V(\pi)$.

Proof for the general case (more than two probabilistic devices) in Subsection A.3.

We continue our running example to show how the previous ambiguous device can be modified.

Example 4. In our running example, the construction of $\overline{V}$ led us to consider the following ambiguous device:

$\pi_1(m\|\omega)$	ω_l	ω_h
m_l	$\frac{2}{3}$	0
m_h	$\frac{1}{3}$	1

$\pi_2(m\|\omega)$	ω_l	ω_h
m_l	$\frac{3}{4}$	$\frac{1}{4}$
m_h	$\frac{1}{4}$	$\frac{3}{4}$

Recall the value of the optimal ambiguous device is $\overline{V}(1/2) = 2/3$. The Sender's expected value of the ambiguous device when computed with regard to π_1 (assuming Receiver reacts to posteriors generated by both devices) is $v_1(p_1, p_2) = 2/3$; while the value when computed with regard to π_2, $v_2(p_1, p_2)$, is however $1/2$. The Sender's maxmin value of the ambiguous device Π is therefore only $1/2$.

However, one can then use corollary 1, using synonyms, in order to increase the value of the ambiguous device to a value arbitrarily close to $2/3$.

To see this, consider the following devices that use duplicated messages m_l and m_h :

$\pi_1(m\|\omega)$	ω_l	ω_h
m_l	$\alpha\frac{2}{3}$	0
m_h	$\alpha\frac{1}{3}$	α
m_l'	$(1-\alpha)\frac{3}{4}$	$(1-\alpha)\frac{1}{4}$
m_h'	$(1-\alpha)\frac{1}{4}$	$(1-\alpha)\frac{3}{4}$

$\pi_2(m\|\omega)$	ω_l	ω_h
m_l	$(1-\alpha)\frac{3}{4}$	$(1-\alpha)\frac{1}{4}$
m_h	$(1-\alpha)\frac{1}{4}$	$(1-\alpha)\frac{3}{4}$
m_l'	$\alpha\frac{2}{3}$	0
m_h'	$\alpha\frac{1}{3}$	α

The devices lead to the following posteriors:

$p_1(\omega\|m)$	ω_l	ω_h
m_l	1	0
m_h	$\frac{1}{4}$	$\frac{3}{4}$
m_l'	$\frac{3}{4}$	$\frac{1}{4}$
m_h'	$\frac{1}{4}$	$\frac{3}{4}$

$p_2(m\|\omega)$	ω_l	ω_h
m_l	$\frac{3}{4}$	$\frac{1}{4}$
m_h	$\frac{1}{4}$	$\frac{3}{4}$
m_l'	1	0
m_h'	$\frac{1}{4}$	$\frac{3}{4}$

with distributions $\tau_1 = (\alpha/3, 2\alpha/3, (1-\alpha)/2, (1-\alpha)/2)$ and $\tau_2 = ((1-\alpha)/2, (1-\alpha)/2, \alpha/3, 2\alpha/3)$. In this case, Receiver will best respond with a_m upon seeing m_l and m_l', which takes place with probability $1/2 + \alpha/6$ under τ^1 and τ^2; and best respond with a_h upon seeing m_h, occuring with probability $\alpha 2/3 + (1-\alpha)1/2$ in both devices.

Hence the value of this ambiguous device is $\alpha 2/3 + (1-\alpha)1/2$. By taking α arbitrarily close to 1, we can get a value arbitrarily close to 2/3.

In the more complicated case where $p_0 = 1/8$, one can also apply corollary 1 in the following manner:

$\pi_1(m\|\omega)$	ω_l	ω_h
m_l	$\alpha\frac{20\varepsilon}{21}$	0
m_h	$\alpha\frac{\varepsilon}{21}$	$\alpha\varepsilon$
m_r	$\alpha(1-\varepsilon)$	$\alpha(1-\varepsilon)$
m_l'	$(1-\alpha)\frac{3}{28}$	$(1-\alpha)\frac{1}{4}$
m_h'	$(1-\alpha)\frac{1}{28}$	$(1-\alpha)\frac{3}{4}$
m_r'	$(1-\alpha)\frac{6}{7}$	0

$\pi_2(m\|\omega)$	ω_l	ω_h
m_l	$(1-\alpha)\frac{3}{28}$	$(1-\alpha)\frac{1}{4}$
m_h	$(1-\alpha)\frac{1}{28}$	$(1-\alpha)\frac{3}{4}$
m_r	$(1-\alpha)\frac{6}{7}$	0
m_l'	$\alpha\frac{20\varepsilon}{21}$	0
m_h'	$\alpha\frac{\varepsilon}{21}$	$\alpha\varepsilon$
m_r'	$\alpha(1-\varepsilon)$	$\alpha(1-\varepsilon)$

This generates the following distributions over posteriors:

messages	$\tau_1(m)$	$p_1(\omega_h\|m)$
m_l	$\frac{5\alpha\varepsilon}{6}$	0
m_h	$\frac{\alpha\varepsilon}{6}$	$\frac{3}{4}$
m_r	$\alpha(1-\varepsilon)$	$\frac{1}{8}$
m'_l	$\frac{1-\alpha}{8}$	$\frac{1}{4}$
m'_h	$\frac{1-\alpha}{8}$	$\frac{3}{4}$
m'_r	$\frac{3(1-\alpha)}{4}$	0

messages	$\tau_2(m)$	$p_2(\omega_h\|m)$
m_l	$\frac{1-\alpha}{8}$	$\frac{1}{4}$
m_h	$\frac{1-\alpha}{8}$	$\frac{3}{4}$
m_r	$\frac{3(1-\alpha)}{4}$	0
m'_l	$\frac{5\alpha\varepsilon}{6}$	0
m'_h	$\frac{\alpha\varepsilon}{6}$	$\frac{3}{4}$
m'_r	$\alpha(1-\varepsilon)$	$\frac{1}{8}$

Again Receiver will choose action a_m upon seeing m_l, m'_l, action a_h upon seeing m_h, m'_h, and action a_l upon seeing m_r, m'_r.

In this case, the value of this ambiguous device is given by $(\alpha\varepsilon\frac{1}{6} - \frac{5}{8}(1-\alpha) - \alpha(1-\varepsilon))$ which converges to $\overline{V}(1/8) = 1/6$ when α and ε converge to 1.

4.3 The necessity of (weak) synonyms

A natural question to ask is whether synonyms are necessary for ambiguous communication device to be strictly beneficial. We will provide some thought in this section.

Fix prior p_0 and communication device (M, Π), we say two messages $m \neq m'$ are *weak synonyms* if they induce the same receiver's best response, i.e., $\hat{a}(P_m) = \hat{a}(P_{m'})$. Note that a communication device that does not use weak synonym will "recommend" a different optimal actions at different messages. This is equivalent to a straightforward "recommendation" device that was used in Kamenica and Gentzkow (2011).

Kamenica and Gentzkow (2011) show that for any Bayesian communication device there is an equivalent straightforward "recommendation" device. In other words, weak synonyms are useless for Bayesian persuasion. The following corollary says is no longer true with ambiguous persuasion. In fact, when $|M| = 2$, any strictly beneficial ambiguous communication device using synonyms cannot be replaced by an equivalent straightforward communication device.

Corollary 2. *For any ambiguous communication device (M, Π) with two messages. If m_1 and m_2 are weak synonyms and $v(p_0, \Pi) > V^B(p_0)$ then there is no equivalent straightforward communication device.*

Proof. Suppose there is an equivalent straightforward communication device. By definition of weak synonym, $\hat{a}(m_1) = \hat{a}(m_2) = \hat{a}^*$ and any equivalent straightforward communication device (M', Π') has only one message $M' = \{\hat{a}^*\}$, which is not informative. So the only Bayes plausible posterior under M' is p_0 and the value of any outcome equivalent straightforward device is $V_0 \leq V^B(p_0)$. These two devices are not payoff equivalent to Sender. $\square$

For example, consider the optimal ambiguous communication device considered in Section 2. Under that device, Receiver's optimal action is to choose $a^* = 0$ (to prescribe brand name drug) upon observing message i and e. Observe that there is no direct "recommendation" mechanism that would generate the same equilibrium outcome. To see this, suppose an equivalent direct "recommendation" mechanism exists, then it must recommend $a^* = 0$ when $m = i$ and $a^* = 0$ when $m = e$. In this case, the recommendation message is not informative and the physician will not update her prior. With prior p_0 the physician will prescribe the generic drug. The physician always recommending brand name drug is not an equilibrium.

Hence with ambiguous messages, focusing on straightforward recommendation device is not without loss of generality. The "revelation principle" à la Kamenica and Gentzkow (2011) may fail under ambiguous communication device. We need to consider the full message space when characterizing the equilibrium. [18]

5 Examples

5.1 When Sender-optimal Action Is Receiver-safe

In the lead example, Sender has an optimal action—prescribing the brandname drug—that is also a safe default for the Receiver. In this case, Sender can benefit from sending two ambiguous messages that are synonyms, as both messages generates the same set of posteriors.

This intuition can be generalized to the case with finitely many states and actions.

We first state a few assumptions to generalize the lead example.

Assumption 1. Sender has a state-independent optimal action, i.e., there is some $a^* \in A$ such that $v(a^*, \omega) \geq v(a, \omega)$ for all $a \in A$ and $\omega \in \Omega$.

Assumption 2. $\min_\omega u(a^*, \omega) \geq \min_\omega u(a, \omega)$.

Assumption 3. Receiver will not choose a^* without persuasion, i.e., there is some $a \in A$ such that $E_{p_0}[u(a, \omega)] > E_{p_0}[u(a^*, \omega)]$.

Assumption 1 clearly holds when Sender's preferences are state-independent, $v(a, \omega) = v(a)$ for all ω. Assumption 2 requires action a^* is also the "safest" for Receiver in the sense that

[18]It worth noting that in the equilibrium we consider Receiver's choice of action still depends solely on the set of posteriors induced by the realized signal. This property is called *language invariant* in Alonso and Câmara (2016), that is, for any two signal structures Π and Π' and realizations s and s' such that $P_m = P_{s'}$, Receiver will choose the same optimal action.

it yields the highest worst-case payoff. To focus on the interesting case, Assumption 3 just requires with only prior information, Receiver will not choose a^*.

Proposition 2. *Under Assumptions 1–2, there exists an optimal ambiguous signal* Π^* *that ensures that Receiver always take the Sender optimal action* a^*.

Under Assumptions 1–3, Sender will do strictly better with Π^* *than with any Bayesian signal.*

Proof. Let $n = |\Omega|$ and pick some $M = \{m_1, m_2, \ldots, m_n\}$. Fix any prior $p_0 \in \Delta(\Omega)$ with full support. Consider the profile of posteriors P_m equals to the extreme points of $\Delta(\Omega)$ for all $m \in M$. By Lemma 1, there exists some set of likelihoods Π that induces these posteriors. By Assumption 2, for all $a \neq a^*$, there is some ω_a such that

$$\min_{\omega} u(a, \omega) \leq u(a, \omega_a) < u(a^*, \omega_a) = \bar{u} = \min_{\omega} u(a^*, \omega),$$

and hence Receiver will choose a^* at all message $m \in M$. Under Assumption 1, this communication device (M, Π) is Sender optimal.

Now, we show that, under Assumption 3, ambiguous messages are also necessary to achieve Sender optimal outcome above. Let $(\tau, \mathbf{p})$ be the Bayes-plausible distribution of posteriors induced by some (unambiguous) Bayesian information structure π^*. Our goal is to show that there always exists a posterior p_m with $\tau(m) > 0$ under which a^* is not an optimal action for Receiver and hence not chosen by her. Suppose instead action a^* is chosen under every posterior p_m such that $\tau(m) > 0$, that is,

$$E_{p_m} u(a^*, \omega) \geq E_{p_m} u(a, \omega) \qquad \text{for all } a.$$

Since $(\tau, \mathbf{p})$ is Bayes-plausible, $p_0 = \sum_s p_s m(s)$. This implies

$$E_{p_0} u(a^*, \omega) = \sum_m \tau(m) E_{p_m} u(a^*, \omega) \geq \sum_m \tau(m) E_{p_m} u(a, \omega) = E_{p_0} u(a, \omega)$$

for all a, contradicting Assumption 3. $\square$

In this example, Sender hedges against any ex ante ambiguity by sending $|M| = |\Omega|$ messages that are synonymous and all generate posteriors with maximal ambiguity, i.e., $P_m = \Delta(\Omega)$, because Receiver will best respond to the same posteriors by taking Sender preferred action a^* at all messages. Sender succeeds in persuasion.

5.2 Crawford and Sobel's (1982) uniform quadratic case

Consider the leading example of Crawford and Sobel (1982), where Sender's and Receiver's payoffs are respectively

$$\begin{aligned} v(a, \omega) &= -(a - \omega)^2, \\ u(a, \omega, b) &= -(a - (\omega + b))^2. \end{aligned}$$

Assume that the random variable ω is uniformly distributed on the interval $[0, 1]$. Let upper case letters U and V denote expected payoffs of Receiver and Sender, respectively.

Applying the Kamenica and Gentzkow (2011) analysis to the above setup, it is straightforward to show that the optimal probabilistic communication device for Sender is to fully disclose ω to Receiver. To see this, let m be the realized message, then Receiver would take an action

$$a^R(m) \quad = \quad E(\omega|m) + b$$

upon observing m. Thus, Sender must choose a communication device to solve the maximization problem

$$\max_{\pi} E\ v\left(a^R(m), \omega\right) = -E\left(a^R(m) - \omega\right)^2.$$

But

$$\begin{aligned} E\left(a^R(m) - \omega\right)^2 &= E\left[E(\omega|m) + b - \omega\right]^2, \\ &= E\left[E(\omega|m) - \omega\right]^2 + 2E\left[E(\omega|m) - \omega\right]b + b^2, \\ &= E\left[E(\omega|m) - \omega\right]^2 + b^2, \end{aligned}$$

which is minimized when

$$E(\omega|m) = \omega.$$

Here, we want to investigate whether it is possible for Sender to do strictly better using ambiguous devices, even if Sender is ambiguity averse himself. We do so through the introduction of ambiguous communication devices with synonymous messages.

Kellner and Le Quement (2016) demonstrate that, in Crawford and Sobel's cheap-talk game, there exist Pareto-improving communication equilibrium when Sender is allowed to send ambiguous messages. Their construction relies on what they call a "simple Ellsberg randomization" device, in which each partition interval, $[\underline{\omega}, \bar{\omega}]$, is associated with two messages that are ambiguous, which could come from either subinterval $[\underline{\omega}, c]$ or $(c, \bar{\omega}]$. These two messages are associated with the same set of posteriors and hence synonymous. Moreover, Receiver's optimal action to each message is the same and determined only by the cutoff c. This helps them to construct more informative equilibria than the standard ambiguity-neutral (Bayesian) setup.

Formally, let $M = \{m_{1A}, m_{1B}, \ldots, m_{nA}, m_{nB}\}$ and $\Pi(\mathbf{y}, \mathbf{c}) = \{\pi, \pi'\}$ be an arbitrary *simple ambiguous communication device* that partitions the $[0, 1]$ interval into $2n$ cells $\{[y_{i-1}, y_{i-1} + c_i], (y_{i-1} + c_i, y_i) : i = 1, \ldots, n\}$ with $y_0 = 0$ and $y_n = 1$. Denote by $I_i = [y_{i-1}, y_i)$, $I_{iA} = [y_{i-1}, y_{i-1} + c_i]$, $I_{iB} = (y_{i-1} + c_i, y_i)$, and $l_i = y_i - y_{i-1}$ for all $i = 1, \ldots, n$.

We consider ambiguous signals denoted by π and π', which satisfy

1. $\pi(\{m_{iA}, m_{iB}\}|\omega) = \pi'(\{m_{iA}, m_{iB}\}|\omega) = \mathbf{1}_{\omega \in I_i}$;

2. $\pi(m_{iA}|\omega) = 1$ if $\omega \in I_{iA}$, and $\pi(m_{iB}|\omega) = 1$ if $\omega \in I_{iB}$;

3. $\pi(m_{iA}|\omega) = 1_{\omega \in I_i} - \pi'(m_{iA}|\omega)$.

Note that communication device π generates m_{iA} if and only if the state is in I_{iA} and m_{iB} if and only if the state is in I_{iB}, while communication device π' does the reverse and generates m_{iA} if and only if the state is in I_{iB} and m_{iB} if and only if the state is in I_{iA}. The induced posteriors for messages m_{iA} and m_{iB} are both $\{Uni(I_{iA}), Uni(I_{iB})\}$, where $Uni(I_{iA})$ refers to the uniform distribution on I_{iA} and vice versa for I_{iB}. In other words, messages m_{iA} and m_{iB} are synonyms. Now, we want to see what Receiver's optimal action is after observing message m_{iA} or m_{iB}.

Observe that the functions u and v are *translation invariant* in (a, ω), in that

$$\begin{aligned} v(a, \omega) &= v(a - t, \omega - t), \\ u(a, \omega, b) &= v(a - t, \omega - t, b) \end{aligned}$$

for all $t \in \mathbb{R}$. Therefore, to simplify notation, we analyze the simple case below and use translation to obtain results for $\{m_{iA}, m_{iB}\}$ and $\{I_{iA}, I_{iB}\}$.

Now consider the interval $I = [0, l)$ and let $I_A = [0, c]$ and $I_B = (c, l)$ for $c \in I$ and let the signals π and π' be analogously defined as above. Now given the set of posteriors $\{I_A, I_B\}$, if Receiver takes an action $x + b$, Receiver's maxmin expected utility can be written (with a slight abuse of notation)

$$\begin{aligned} U(x+b) &= \min\left\{E_{[0,c]}u(x+b, \omega, b), E_{(c,l)}u(x+b, \omega, b)\right\}, \\ &= \min\left\{-x^2 + 2xE_{[0,c]}\omega - E_{[0,c]}\omega^2, -x^2 + 2xE_{(c,l)}\omega - E_{(c,l)}\omega^2\right\}, \\ &\equiv \min\left\{h_1(x), h_2(x)\right\}. \end{aligned}$$

Note that h_1 and h_2 are both concave quadratic functions with maxima at $c/2$ and $(c+l)/2$ respectively. In addition, $h_1(x) \leq h_2(x)$ if and only if

$$2xE_{[0,c]}\omega - E_{[0,c]}\omega^2 \quad \leq \quad 2xE_{[c,1]}\omega - E_{[c,1]}\omega^2, \tag{3}$$

which simplifies into

$$x \quad \geq \quad \frac{l+c}{3} \equiv x^*. \tag{4}$$

Note that x^* is clearly in $(c/2, (l+c)/2)$, i.e., between the maximum points for h_1 and h_2. By properties of quadratic functions, we know

- Receiver's expected payoff $Eu(x+b) = h_2(x)$ for $x \leq x^*$ and $Eu(x+b) = h_1(x)$ for $x \geq x^*$.

- The function h_1 is decreasing to the right of x^* and h_2 increasing to the left of x^*.

From these statements, we may conclude $U(x+b)$ increases up to x^* and then decreases. Therefore, it reaches its maximum at x^*. So, we conclude that Receiver's optimal action when observing m_A is

$$x^* + b = \frac{l+c}{3} + b. \tag{5}$$

Sender's ex-ante utility is

$$V(\Pi(c), b) = -E_{p_0}\left(\frac{l+c}{3} + b - \omega\right)^2.$$

Since $V(\Pi(c), b)$ is a concave function of c, the interior sender optimal cutoff $c^*(b)$ is determined by the first order condition

$$\begin{aligned}
\frac{\partial V(\Pi)}{\partial c} &= -\frac{\partial}{\partial c}E_{[0,l)}[\frac{c}{3} + b + \frac{l}{3} - \omega]^2 = 0 \\
&\Rightarrow -E_{[0,l)}[\frac{2c}{9} + \frac{2}{3}[(b + \frac{l}{3} - \omega)] = 0 \\
&\Rightarrow \frac{c}{3} + \left(b + \frac{l}{3} - E_{[0,l)}[\omega]\right) = 0 \\
&\Rightarrow c = \frac{l}{2} - 3b
\end{aligned}$$

Combined with the domain restriction $c \in [0, l)$, we have

$$c^*(l, b) = \begin{cases} \frac{l}{2} - 3b, & \text{if } b < \frac{l}{6}; \\ 0, & \text{if } b \geq \frac{l}{6}, \end{cases}$$

and

$$a^*(l) = \begin{cases} a_1(l) \equiv \frac{l}{3} + b, & \text{if } l \leq 6b; \\ a_2(l) \equiv \frac{l}{2}, & \text{if } l \geq 6b. \end{cases}$$

We say a simple ambiguous communication device $\Pi(\mathbf{y}, \mathbf{c}) = \{\pi, \pi'\}$ is *symmetric* if $\mathbf{y} = (0, 1/n, \ldots, (n-1)/n, 1)$ and $\mathbf{c} = (c, \ldots, c)$ for some $n \in \mathbb{N}$ and $c \in [0, 1/n)$. We use $\Pi(1/n, c)$ to denote such a symmetric ambiguous communication device.

Lemma 3. *For all simple ambiguous communication device* $\Pi(\boldsymbol{y}, \boldsymbol{c})$*, there exists a symmetric simple ambiguous communication device* $\Pi(1/n, c)$ *such that* $V(\Pi(1/n, c)) \geq V(\Pi(\boldsymbol{y}, \boldsymbol{c}))$.

The above lemma implies that we may without loss of generality focus on ambiguous communication devices that are symmetric. The induced posteriors for messages m_{iA}, m_{iB} are $\{[(i-1)/n, (i-1)/n + c_i], ((i-1)/n + c_i, i/n)\}$, where $[(i-1)/n, (i-1)/n + c_i]$ refers to the uniform distribution on $[(i-1)/n, (i-1)/n + c_i]$, and vice versa for $((i-1)/n + c_i, i/n)$. Again, this can be generated by a pair of messaging technologies as described above.

Proposition 3. *In the uniform-quadratic case, considering the set of simple ambiguous communication devices,*

1. *Sender always benefits from sending ambiguous messages;*
2. *there exists an $n^*(b)$, such that among the simple ambiguous communication devices, $\Pi(1/n^*(b), 0)$ achieves the highest payoff for Sender;*
3. *Receiver's participation constraint is satisfied if and only if the number of intervals is greater than or equal to* 2.

Kellner and Le Quement (2016) (KLQ hereafter) characterize two classes of special equilibria: equal intervals equilibria (see Proposition 4 of their paper) and maximum ambiguity equilibria (see Proposition 5 of their paper). Equal intervals refers to the fact that the state space $[0, 1]$ is divided into equal-length intervals as a first step, and each interval is assigned a cutoff for constructing the simple ambiguous communication device. Maximum ambiguity refers to the fact that the cutoff chosen for each interval is at the endpoint of the interval. In KLQ's model, these two properties cannot be jointly satisfied, given that $b \neq 0$. Given that the IC constraint of Sender has to be respected, Receiver must choose an action equal to the midpoint of each interval, behaving as if she had no bias relative to Sender.[19]

Thus, according to our (and KLQ's) characterization of Receiver's best response in (5), it must be that

$$\frac{l+c}{3} + b \quad = \quad \frac{l}{2},$$

or

$$c \quad = \quad \frac{l}{2} - 3b.$$

This has two implications. First, the length of each interval must be at least equal to $6b$. Second, unless $l = 6b$, the equal length equilibrium does not satisfy maximum ambiguity. The first implication in turn requires that, in order for there be at least two intervals in an equilibrium, b must be lower than $1/12$, which is a more stringent requirement than the Crawford-Sobel threshold for informative equilibria, $1/4$. In a maximum ambiguity equilibrium, the cutoffs can be obtained by using (5) and Sender's incentive compatibility constraint. Let us focus on two adjacent intervals of length l and l', we must have

$$l - \left(\frac{l+c}{3} + b\right) \quad = \quad \left[l + \left(\frac{l'+c'}{3} + b\right)\right] - l,$$

[19]For expositional convenience, in this paper we employ a setup in which Receiver has a positive bias relative to Sender. Our discussion of KLQ's results here also employs our setup rather than theirs, where Sender has a positive bias relative to Receiver. However, our setup is equivalent to a setup in which Sender has a negative bias relative to Receiver. In our setup, each equilibrium in the communication game is the mirror image of an equilibrium when Sender has a positive bias relative to Receiver. The unique equilibrium we characterize under ambiguous persuasion is also the mirror image of that when Sender has a positive bias relative to Receiver.

where $c = c' = 0$. The equation simplifies into

$$\begin{aligned} l - \left(\frac{l}{3} + b\right) &= \left(\frac{l'}{3} + b\right), \\ l' &= 2l - 6b. \end{aligned}$$

Again, this demonstrates that a maximum ambiguity equilibrium can be equal length if and only if each interval is of length $6b$. The above equation defines all the cutoff partitions recursively, analogous to what Crawford and Sobel (1982) do. The equilibrium has the maximum number of intervals is when the shortest interval in the partition has zero length.

In contrast, the optimal information structure we characterize in Proposition 3 of Section 5.2 satisfies both equal intervals and maximum ambiguity. Given our assumption that Sender has full commitment, no incentive compatibility requirement is placed on Sender. The payoff that is achievable by Sender is therefore much improved. In our characterized optimal information structure, the state space $[0, 1]$ is divided into $n^*(b)$ equal-length intervals, and then within each interval, Sender creates maximum ambiguity by setting the cutoff at the leftmost point of that interval. By so doing, Sender is able to induce Receiver to take an action that is closer to his most preferred action for each interval. Our characterization demonstrates that endogenous ambiguity serves a purpose for Sender. Without such possibility, Sender would simply completely reveal all the information to Receiver. However, with such possibility, Sender faces a tradeoff between being specific and being ambiguous, the former to reduce the likelihood that decisions made by Receiver are too far away from the state of the world, and the latter to take advantage of Receiver's ambiguity aversion to sway her decision towards his ideal action.

6 Discussion

6.1 Dynamic Consistency

Ambiguity averse decision makers may be dynamically inconsistent. In this section, we consider the set of ambiguous communication devices that induce consistent behaviour from any decision maker. I show that such a set, though not empty, does not allow the expert to benefit from ambiguous communication. I say that an ambiguous device is dynamically consistent if *any* decision maker would play in a dynamically consistent manner when receiving messages from such an ambiguous device.

For this purpose, let $U(a, \Pi) = \min_{\pi \in \Pi} \sum_\omega p_0(\omega) \left[\sum_m \pi(m|\omega)\ u(a_m, \omega)\right]$ be the *ex-ante* MaxMin Expected Utility of the receiver when he plays strategy $a \in A^M$.

Definition 1. Π is dynamically consistent if for all A and $u : A \times \Omega \to \mathbb{R}$ and for all m, $U(a_1, P_m) \geq U(a_2, P_m) \Rightarrow U(a_1, \Pi) \geq U(a_2, \Pi)$.

In a more general model, Epstein and Schneider (2003) show that dynamic consistency is equivalent to rectangularity of the priors. Note that rectangularity is defined in their paper over the full state space[20] $\Omega \times M$. We adapt here their definition to this paper's framework. Let $\hat{P}_0 = \{p | p(\omega, m) = p_0(\omega)\pi(m|\omega) \text{ for } \pi \in \Pi\}$ be the set of priors in the full state space and $\hat{P}_m = \{p | p(\omega, m) = p_\pi(\omega) \text{ and } p(\omega, m' \neq m) = 0 \text{ for } \pi \in \Pi\}$ the set of posteriors in the full state space. The definition of rectangularity from Epstein and Schneider (2003) is that $\hat{P}_0 = \{\sum_m q(m)p_m | q \in P_0, (p_m)_M \in \hat{P}_m\}$. This is equivalent to $\{p_0\} = \{\sum_m \pi(m)p_{\pi_m}(\cdot|m) | (\pi, (\pi_m)_M) \in \Pi^{M+1}\}$ in the restricted state space used here.

Definition 2. Π is said rectangular with respect to p_0 if for all $(\pi, (\pi_m)_M) \in \Pi^{M+1}$, $p_0 = \sum_m \pi(m)p_{\pi_m}$.

Rectangularity makes sure the set of priors (in the full state space) is "complete" in the sense that any profile of posteriors can be obtained from the set of priors, with any profile of weights possible. Note that P_0 in this case must necessarily be the singleton $\{p_0\}$ so that the definition of rectangularity collapses to:

Lemma 4. Π *is rectangular if and only if* P_m *is a singleton for all* m.

Proof. We only prove the direct implication here. Assume by contraposition that there exists m such that $p_\pi(\cdot|m) \neq p_{\pi'}(\cdot|m)$, then $p = \pi(m) \cdot p_{\pi'}(\cdot|m) + \sum_{m' \neq m} \pi(m') \cdot p_\pi(\cdot|m')$ is different from p_0. □

As in Epstein and Schneider (2003), rectangularity is, here, equivalent to dynamic consistency. The following result is a Corollary of their representation theorem.

Corollary 3. Π *is dynamically consistent if and only if* Π *is rectangular.*

Proof in Subsection A.7.

If the expert is restricted to dynamically consistent devices, then, from Lemma 4, he may use only rectangular devices. Note that rectangular devices are not equivalent to probabilistic devices. Indeed, if a rectangular device uses more messages than there are states of the world, there is some more freedom to choose different weights from one probabilistic device

[20] Rectangularity must also be defined for a given filtration. In this case, the filtration is naturally $\{\Omega \times \{m\}\}_M$

to another. Nevertheless, any action that such a device would entail could have been induced by a risky device as well. Unsurprisingly then, the expert cannot benefit from ambiguous persuasion if he is restricted to rectangular devices.

Proposition 4. *If the expert is restricted to dynamically consistent/rectangular strategies, then there are no gains to playing ambiguous strategies.*

Proof. Let $\pi \in \arg\min_{\pi \in \Pi^*} V(\pi, a^*)$. Then playing the mixed strategy π yields the same strategy of the receiver (given rectangularity implies singleton posteriors) and thus $V(\{\pi\}) = V(\Pi^*)$. $\square$

Note that the restriction to dynamically consistent devices is a strong restriction. First, it imposes a constraint on the receiver's preferences over actions which he does not take in equilibrium. Second, it considers not only that the strategy of the receiver facing the expert must be dynamically consistent but also that the strategies of *any* receiver be dynamically consistent.

Weakening any of these two conditions would allow the expert to benefit from ambiguous communication. The next subsection relaxes the first condition while the following relaxes the second.

6.2 Weak dynamic consistency

In this section, we look at a weaker condition, which only requires the receiver's preferences over actions to be consistent on the equilirbium path.

Fix some ambiguous device Π, it induces a set of distributions over posterior R that is generalized Bayes plausible. Let $(P_m)_{m \in M}$ be the profile of posterior sets projected from R to the restricted domain $(\Delta\Omega)^M$ and $a_m^* = \hat{a}(P_m)$ be the (sender-preferred) optimal action by Receiver at posterior set P_m. Denote by Q_m^* the set of posteriors from P_m that supports the optimality of a_m^*, i.e.,

$$Q_m^* := \{q_m \in co(P_m) : E_{q_m}[u(a_m^*, \omega)] \geq E_{q_m}[u(a, \omega)] \qquad \forall a \in A\}.$$

In words, Q_m^* is the set of posteriors at which an SEU Receiver with some equivalent posterior $q_m^* \in Q_m^*$ would choose the same optimal action as an MEU Receiver with posteriors P_m.

Example 5. In our lead example, the set of posteriors are $P_m = co\{(0, 1), (1, 0)\}$ and Receiver's best response is to choose brandname at both messages. Hence the set of posteriors supporting a_m^* =brandname is $Q_m^* = \{(p, 1 - p) : p \in [0, 1 - \frac{c}{u_H - u_L}]\}$ for $m = e, i$.

The consistency closure of R is $\bar{R} = \{(\tau, \mathbf{p}) \in \Delta M \times (\Delta\Omega)^M : \mathbf{p} \in (P_m)_{m\in M}, \sum_m \tau(m)p_m = p_0\}$.

Definition 3. The ambiguous device Π is *semi-rectangular* if its induced set of distributions over posterior satisfies $R = \bar{R}$.

The next proposition provides a necessary condition for some ambiguous message inducing R to be strictly beneficial for Sender. Intuitively, there must not be an "equivalent" Bayes plausible distribution of posterior that induces the same Receiver best response.

The next proposition says Sender can never strictly gain from a semi-rectangular ambiguous device.

Proposition 5. *If an ambiguous communication device (M, Π) is semi-rectangular and the profile of supporting posteriors $(Q^*_m)_{m\in M}$ is nonempty and potentially generalized Bayes plausible, then there is some Bayesian communication device that is weakly better than (M, Π).*

Proof in Subsection A.8.

Note that semi-rectangularity requires the set of induced distributions of posteriors R to be sufficiently large so that it includes all plausible mixtures of posteriors. The condition that $(Q^*_m)_{m\in M}$ is potentially generalized Bayes plausible is more substantial for an ambiguous device to be strictly beneficial. In our lead example, $Q^*_m = \{(p, 1-p) : p \in [0, 1 - \frac{c}{u_H - u_L}]\}$ for $m = e, i$ but $1 - \frac{c}{u_H - u_L} < p_0$. Hence $(Q^*_m)_{m\in M}$ is not potentially generalized Bayes plausible, which is essential for the described ambiguous device to be strictly beneficial than Bayesian devices.

6.3 Positive Value of Information

In this section, we consider a weaker condition than dynamic consistency on Π, namely that the value of information must be positive. Whereas before we asked that each strategy profile be ranked in the same order ex ante and ex post, here we ask only that their ranking with the default action be the same. In other words, *all* receivers must benefit from the ambiguous device.

Definition 1. Π *is* valuable *(to the receiver) if for all utility function $u : A \times \Omega \to \mathbb{R}$, $U(\Pi) \geq U_0$*

Schlee (1997) show that valuableness is equivalent to dynamic consistency when payoffs may depend on the full state space. In this paper however, payoffs must be constant on the message dimension of the full state space. This restriction on the framework breaks the equivalence between dynamic consistency and valuableness. As a result, it is possible to benefit from ambiguous communication that would be valuable to any receiver.

Consider for example the following payoffs :

	ω_l $(\frac{1}{2})$	ω_h $(\frac{1}{2})$
a_l	$-1; 1$	$-1; -2$
a_m	$0; 0$	$0; 0$
a_h	$1; -2$	$1; 1$

There are two optimal probabilistic devices here. Note that, as in the previous examples, probabilities refer to the probability that the high state occurs. The first one, denoted π, yields the posteriors 0 at the low message, with probability 1/4, and 2/3 at the high message, with probability 3/4. The second one, denoted π_2, yields the posteriors 1/3 at the low message and 2/3 at the high message with equal probability. In both cases, $V(\pi_1) = V(\pi_2) = 1/2$.

We now construct an ambiguous device Π from which the expert would benefit as in section 4 using the probabilistic device π_1 which yields the posteriors 1/6 at the low message, with probability 1/3, and 2/3 at the high message, with probability 2/3. Now consider the ambiguous device $\Pi = co((\pi_1, \pi_2))$. The receiver plays the middle action at the low message. The value of this device, when computed with regard to π_1 is $v_1(\Pi) = 2/3$. Thus the expert could benefit from ambiguous communication by using synonyms. Before doing so however, we transform Π into a valuable communication device.

Consider now the ambiguous device $\Pi' = co(\frac{2}{3}\pi \oplus \frac{1}{3}\pi_1; \frac{2}{3}\pi \oplus \frac{1}{3}\pi_2)$. This device leads to the posterior sets $\{0\}$, $[1/6; 1/3]$ and $\{2/3\}$. Furthermore, this device can be shown to be valuable. The key idea is that any loss of utility from ambiguous information (at the $[1/6; 1/3]$ posterior) would be partly offset by the gains at the $\{0\}$ posterior. If the message (at 0) is sent sufficiently often compared to the ambiguous message, the gains can be shown to always offset the losses.

For a receiver to lose utility at the ambiguous message, it must be that he chooses an action which is different from the default at that message. Furthermore, this action may not strictly dominate the default action at that posterior. In the worst case, the default action and the chosen action are equivalent at the 1/6 posterior. The maximum loss possible is therefore equal to $a|1/3 - 1/6|$ where a is the difference of slope between the payoffs of both actions. On the other hand, the gains of playing the new action instead of the default one at posterior 0 is therefore $a|1/6-0|$. The expected gains and losses from this ambiguous device are then at worst $a\pi(\{0\})|1/6-0| - a\pi([1/6, 1/3])|1/3-1/6|$ where π must be the probabilistic device that yields the posterior 1/3. Indeed, under the other probabilistic device, both the default action and the chosen action are equivalent at the induced posterior of 1/6. In our example, this gives $a(1/6*2/3*1/4-1/6*1/2*1/3)=0$. This explains why in the construction of Π', a weight larger than 1/3 was not attributed to π_1 and also why it was not possible to choose a device π_1 which would have led to a posterior of much less than 1/6.

Subsection A.9 proves more rigorously a (slightly) more general result[21].

Note that although Π' is valuable, it is not proven yet that the expert benefits from it. Indeed, we have for the moment that $V(\Pi') = \min(2/3*1/2+1/3*2/3; 2/3*1/2+1/3*1/2) = 1/2$. Nevertheless, given the (ex ante) utilities of the receiver and the sender are both linear in the $\oplus$ operation, it follows that valuableness is conserved when using synonyms as in corollary 1. Thus, the expert can use synonyms to capture, with a valuable ambiguous device, the value of $5/9 > 1/2$.

6.4 Participation Constraint

In this section, we relax the condition that an ambiguous device must be satisfactory for all receivers. We present two methods which can allow the expert to modify an ambiguous device in order to satisfy the following participation constraint: $\sum \tau_m^k \tilde{u}_k(P_m) \geq U_0$ in R.

One first solution is to mix (in the $\oplus$ sense) the optimal ambiguous device (without participation constraint) with a probabilistic one in the same manner as the synonyms were created. This method should be reminiscent of the previous section. However, the valuable condition above restricted the set of ambiguous devices one could use. Indeed, in the above example, it would not have been possible to have the ambiguous posterior to have the value $[0; 1/3]$. The following result lifts this constraint.

Proposition 6. *If the expert benefits from ambiguous communication (if $\bar{V}(p_0) > \hat{V}(p_0)$) and the decision maker benefits from the optimal Bayesian communication, then both the expert and the decision maker benefit from ambiguous communication.*

[21] We have here that $\underline{p} = 1/6$, $\overline{p} = 1/3$, $p_l = 0$ and $\overline{\pi}(m) = 1/2*1/3$ and $\overline{\pi}(l) = 2/3*1/4$.

Proof. Let Π be the optimal ambiguous device and π be the optimal risky device. Let $\Pi' = \alpha\Pi \oplus (1-\alpha)\pi$ be the ambiguous device where each of its probabilistic devices are mixed with the optimal risky one. The value of this new device is necessarily greater than the value of the optimal risky device. Thus, for all $\alpha > 0$, $V(\Pi') > V(\pi)$. Furthermore, the receiver's value of said device is $U(\Pi') = \alpha U(\Pi) + (1-\alpha)U(\pi)$. Given $U(\pi) > U_0$ by assumption, it is always possible to find $\alpha > 0$ such that $U(\Pi') \geq U_0$.

□

Consider for example the game represented by the following Figure 5 that provides the $\hat{v}$ function.

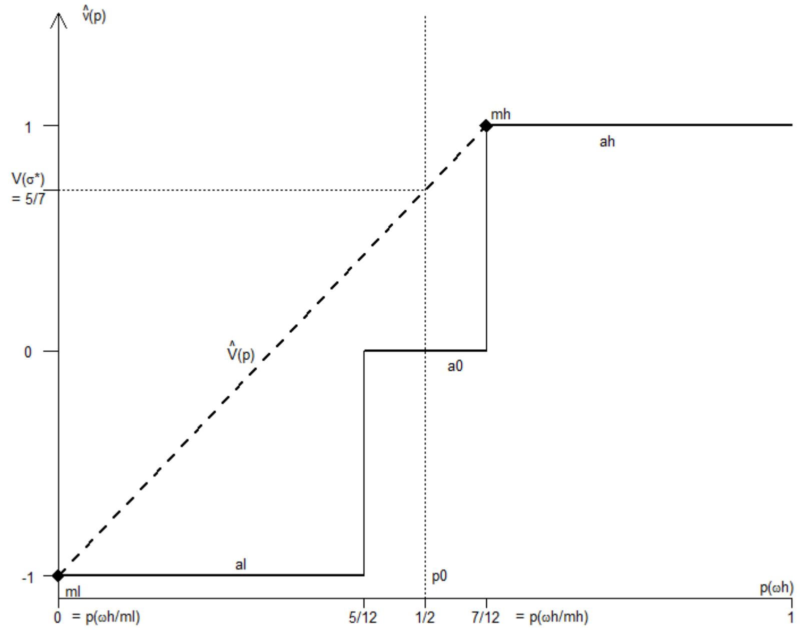

Figure 5: The value of ambiguous communication

In this case, the decision maker would benefit strictly from probabilistic communication. Thus, if it happens that the expert benefits from ambiguous communication[22], it is possible to find an ambiguous device such that both the decision maker and the expert benefit.

A second method, which is illustrated by the second example, is to restrict the expert to

[22]If the high action is the safest one for the decision maker, then the first example's method would work. If the default action is the safest one, then the second example's method would work as well. The only case where ambiguous communication would have no benefit here is if the safest action was the low action.

"value-increasing" messages.

Let $p^{-1}(a_0) = \{p \in \Delta(\Omega)/\forall a \in A, u(a_0, p) > u(a, p)\}$ be the set of priors under which the default action is strictly preferred to other actions.

A message is value-increasing (to the decision maker) if $u_i(\hat{a}(P_m), p_m) \geq u(a_0, p_m)$ for all posteriors in P_m. The following characterization applies:

Proposition 1. *A message is value-increasing (to the receiver) if and only if $P_m \cap p^{-1}(a_0) = \emptyset$ or $\hat{a}(P_m) = a_0$. Additionally, if Π uses only value-increasing messages, then the receiver benefits from Π.*

Proof in section A.10.

The main interest of this characterization is that it allows us to assign a weight of $-\infty$ to $\tilde{v}_k(P_m)$ if P_m is not a value-increasing message. Thus, using the characterizations from the Propositions and Lemmas of sections 4 and 4.2 directly would ensure that the resulting ambiguous device satisfies the participation constraint of the receiver.

This second proposition is also of interest as it is possible that the receiver plays in a dynamically consistent manner. Indeed, this is the case for the example presented in section 4. In order to ensure dynamic consistency in a given game, one could simply use the restriction that the receiver holds a weakly dominant action at each posterior set: $P_m \cap p^{-1}(a) = \emptyset$ or $\sigma(P_m) = a$ for all actions $a \in A$.

Furthermore, using this method and dropping valuableness, enables the expert to get a payoff of 3/4 in the game presented in section 6.3 instead of 5/9.

7 Conclusion

In this paper, we showed that an informed expert could often beneficially introduce ambiguity in his communication strategy, when agents are ambiguity averse à la Gilboa and Schmeidler (1989). We view this as a justification for why in a variety of situations of persuasion, the information provided by expert is ambiguous and subject to multiple interpretations. One of the interesting features of ambiguous communication is the fact that the expert will use synonyms in order to hedge himself against ambiguity. This remark therefore provides a simple way of differentiating ambiguous communication from Bayesian communication.

In addition, this insight is robust to a dynamic consistency requirement, so long as it is not required for all receivers and a valuableness requirement.

Finally, we explore applications in two examples. In the first one, the expert can deliberately use ambiguity to muddy the waters and persuade the receiver into taking a safe action. In the second one, in a Crawford and Sobel (1982) "uniform-quadratic" setting, a biased expert can achieve strictly higher payoff by sending imprecise information enhanced by ambiguity, so as to influence the receiver to take actions that are more in line with the expert's ideal actions.

A Appendix: Proofs

A.1 Proof of Lemma 1

Proof. (1)$\Rightarrow$ (2). Pick an arbitrary $p_m \in P_m$, by statement (1) there exists some $\pi \in \Pi$ such that $p_m = p^{\pi}_{0,m}$. For all $m' \in M \backslash m$, define $p'_{m'} := p^{\pi}_{0,m'}$. By statement (1), $p'_{m'} \in P_{m'}$ for all $m' \in M \backslash s$. Finally, let $p'_m = p_m$ and construct the posterior profile $(p'_{m'})_{m' \in M} \in (P_{m'})_{m' \in M}$. It is easy to verify properties (i) and (ii) in statement (2). Let $\tau_{m'} = \int_{\Omega} \pi(m'/\omega')p_0(\omega')d\omega'$, then it is straightforward to verify that $\sum_{m'} \tau_{m'} = 1$ and

$$\begin{aligned}\sum_{m'} \tau_{m'} \frac{p'_{m'}(\omega)}{p_0(\omega)} &= \sum_{m'} \int_{\Omega} [\pi(m'/\omega')p_0(\omega')d\omega'] \frac{p'_{m'}(\omega)}{p_0(\omega)} \\ &= \sum_{m'} \int_{\Omega} [\pi(m'/\omega')p_0(\omega')d\omega'] \frac{\pi(m'/\omega)p_0(\omega)}{\int_{\Omega} \pi(m'/\omega')p_0(\omega')d\omega'} \frac{1}{p_0(\omega)} = 1\end{aligned}$$

for all ω.

(2) $\Rightarrow$ (1). By statement (2), for all $m \in M$ and $p_m \in P_m$, there exists a profile of posteriors $(p'_{m'})_{m' \in M} \in (P_{m'})_{m' \in M}$ that satisfies properties (i) and (ii). This posterior profile induces a signalling policy $\pi(m'/\omega) := \frac{p'_{m'}(\omega)\tau_{m'}}{p_0(\omega)}$. Let Π be the set of signalling policies induced by all $m \in M$ and $p_m \in P_m$.

Then for all $m \in M$ and $\pi \in \Pi$, then it is induced by a profile of posteriors $(p'_{m'})_{m' \in M} \in$

$(P_{m'})_{m'\in M}$ that satisfies properties (i) and (ii). For all ω,

$$\begin{aligned} p^{\pi}_{0,m}(\omega) &= \frac{\pi(m/\omega)p_0(\omega)}{\int_{\Omega'}\pi(m/\omega')p_0(\omega')d\omega'} = \frac{p_0(\omega)\pi(m/\omega)}{\tau_m} \\ &= \frac{p_0(\omega)\tau_m p'_m(\omega)}{p_0(\omega)\tau_m} = p'_m(\omega) \end{aligned}$$

Hence $\cup_{\pi\in\Pi}\{p^{\pi}_{0,m}\} \subseteq P_m$.

For all $m \in M$ and $p_m \in P_m$, by statement (2) there exists a profile of posteriors $(p'_{m'})_{m'\in M} \in (P_{m'})_{m'\in M}$ that satisfies properties (i) and (ii). By construction there is some likelihood function $\pi \in \Pi$ such that $\pi(m'/\omega) = \frac{p'_{m'}(\omega)\tau_{m'}}{p_0(\omega)}$. For all ω,

$$p_m(\omega) = \frac{p_m(\omega)\tau_m}{\tau_m} = \frac{p_0(\omega)\pi(m/\omega)}{\int_{\Omega'}\pi(m/\omega')p_0(\omega')d\omega'} = p^{\pi}_{0,m}(\omega)$$

Hence $P_m \subseteq \cup_{\pi\in\Pi}\{p^{\pi}_{0,m}\}$. □

A.2 Proof of Lemma 2

Proof. Let $\pi' = \alpha\pi^1 \oplus (1-\alpha)\pi^2$. In order to avoid confusion, I denote here $a^*_m(\pi)$ the strategy of the receiver when he receives message s, which was sent through the π bayesian communication device.

Note that $a^*_{m_1}(\pi') = a^*_{m_1}(\pi^1)$ as the posterior belief at m_1 of the receiver when π' is the bayesian device is given by $\frac{p_0(.)\pi'(m_1/.)}{\pi'(m_1)} = \frac{p_0(.)\alpha\pi^1(m_1/.)}{\alpha\pi^1(m_1)} = \frac{p_0(.)\pi^1(m_1/.)}{\pi'^1(m_1)}$ and is therefore equal to the posterior belief at m_1 when π^1 is the bayesian device. Likewise, $a^*_{b(m_1)}(\pi') = a^*_{m_1}(\pi^2)$.

$$
\begin{aligned}
V(\pi') = & \sum_{\omega} p_0(\omega)\left[\sum_{m\in M} \pi'(m/\omega)\ v(a^*_m(\pi'),\omega)\right] \\
= & \sum_{\omega} p_0(\omega)\left[\sum_{m_1\in M_1} \pi'(m_1/\omega)\ v(a^*_{m_1}(\pi'),\omega)\ + \sum_{m_2\in M_2} \pi'(m_2/\omega)\ v(a^*_{m_2}(\pi'),\omega)\right] \\
= & \sum_{\omega} p_0(\omega)\left[\sum_{m_1\in M_1} \alpha\pi^1(m_1/\omega)\ v(a^*_{m_1}(\pi'),\omega)\ + \sum_{b(m_1)\in M_2} (1-\alpha)\pi_2(m_1/\omega)v(a^*_{b(m_1)}(\pi'),\omega)\right] \\
= & \sum_{\omega} p_0(\omega)\left[\sum_{m_1\in M_1} \alpha\pi^1(m_1/\omega)\ v(a^*_{m_1}(\pi^1),\omega)\ + \sum_{m_1\in M_1} (1-\alpha)\pi^2(m_1/\omega)v(a^*_{m_1}(\pi^2),\omega)\right] \\
= & \ \alpha\left[\sum_{\omega} p_0(\omega)\left(\sum_{m_1\in M_1} \pi^1(m_1/\omega)\ v(a^*_{m_1}(\pi^1),\omega)\right)\right] \\
& \qquad + (1-\alpha)\left[\sum_{\omega} p_0(\omega)\left(\sum_{m_1\in M_1} \pi^2(m_1/\omega)v(a^*_{m_1}(\pi^2),\omega)\right)\right] \\
= & \ \alpha V(\pi_1) + (1-\alpha)V(\pi_2)
\end{aligned}
$$

□

A.3 Proof of Corollary 1

Proof. First, I define the $\oplus$ operation for more than two devices. Given a finite family of probabilistic devices $(\pi^k)_K$ that all use the same messages in M_1, let M_k for $k \in K\backslash\{k\}$ a series of sets of messages duplicated from M_1: there exists $K-1$ permutations b_k from M_1 to M_k.

Given a probability function λ over K, I denote $\pi' = \lambda(1)\pi^1 \oplus \lambda(2)\pi^2 \oplus ... \oplus \lambda(K)\pi^K = \bigoplus_k \lambda(k)\pi^k$ as the device which sends a message m_1 from M_1 with probability $\lambda(1)\pi^1(m/\omega)$ from state ω and messages $m_k = b_k(m_1) \in M_k$ with probability $\lambda(k)\pi^k(m_1/\omega)$ from state ω.

Let $\Pi = co((\pi_k)_K)$. Assume without loss of generality that $V(\pi^1) = \sup_k V(\Pi)$. Let γ_t be the permutation such that $\gamma_t(k) = k+t$ if $k+t \leq K$ and $\gamma_t(k) = k+t-K$ otherwise: γ_t is the permutation that increments the indices by t.

For some value $\alpha \in [0,1]$, define π'^k be the probabilistic device defined by $\bigoplus_k \lambda(\gamma_{k-1}(k))\pi_{\gamma_{k-1}(k)}$ where $\lambda(j) = \alpha$ if $j = 1$ and $\lambda(j) = (1 - \frac{\alpha}{K-1})$ otherwise.

Let $\Pi' = co((\pi'_k)_K$ be the ambiguous device that uses messages in $\cup_k M_k$. Below is this device in matrix form:

Π'	π'^1	π'^2	...	π'^K
$m_1 \in M_1$	$\alpha\pi^1(m_1/\omega)$	$(1-\frac{\alpha}{K-1})\pi^2(m_1/\omega)$	...	$(1-\frac{\alpha}{K-1})\pi^K(m_1/\omega)$
$b_2(m_1) \in M_2$	$(1-\frac{\alpha}{K-1})\pi^2(m_1/\omega)$	$(1-\frac{\alpha}{K-1})\pi^3(m_1/\omega)$	...	$\alpha\pi^1(m_1/\omega)$
...	...	...	...	...
$b_K(m_1) \in M_K$	$(1-\frac{\alpha}{K-1})\pi^K(m_1/\omega)$	$\alpha\pi^1(m_1/\omega)$	...	$(1-\frac{\alpha}{K-1})\pi^{K-1}(m_1/\omega)$

From lemma 2, the value of device π'^k is $\alpha V(\pi^1) + (1-\frac{\alpha}{K-1})\sum_{k\neq 1} V(\pi^k)$ for all $k \in K$. As a result, $V(\Pi') = \alpha V(\pi^1) + (1-\frac{\alpha}{K-1})\sum_{k\neq 1} V(\pi^k)$.

Thus, $\underset{\alpha\to 1}{lim} V(\Pi') = \sup_k V(\pi^k) = \sup_{\pi\in\Pi} V(\pi)$.

$\square$

A.4 Proof of Proposition 1

Proof. **If.** To prove that the sender benefits if $\overline{V}(p_0) > \hat{V}(p_0)$, we show that, for arbitrarily small ϵ, there exists an ambiguous device Π such that $V(\Pi) = \overline{V}(p_0) - \epsilon$.

Let $\tilde{P^{-1}} = co(\tilde{p}_k)_{k\geq 2} \in \underset{P^{-1}\in\Delta(\Omega)^{I-1}}{argmax} V_1(\{p_0, P^{-1}\})$.

By construction of V_1, there exists τ_m and p^1_m such that:

$$\sum \tau_m p^1_m = p_0$$
$$\forall k \geq 2, \sum \tau_m p^k_m = \tilde{p}_k$$

By construction, $\{p^1_m\}$ are Bayes plausible with respect to p_0. This implies there exists a bayesian device π_1 that leads to those posteriors.

Consider now the posterior sets $(\tilde{P}^{-1}_m)$. Assume first these posterior sets are potentially generally Bayes plausible. In this case, from Lemma 1, there exists a set of $K-1$ bayesian devices $(\pi_k)_{k\geq 2}$ that lead to these posterior sets. As a result, the ambiguous device defined by $co(\pi_1, \pi_{k\geq 2})$ is an ambiguous device whose value for the sender, when computed with regard to π_1 is equal to $\overline{V}(p_0)$.

Assume now that the posteriors $(\tilde{P}_m^{-1})_{m\in M}$ are not potentially generalized Bayes plausible. In this case, it is possible to extend these profile of posterior sets so that they would be. Construct the posterior set in the following manner: for each $k \geq 2$, let $p^k_{m_r}$ be a distribution such that $p_0 \in co(p^k_{m_r}, (p^k_m)_{m\in M})$ as long as p_0 is of full support in $\Delta(\Omega)$. Otherwise, then one would restate the problem to only those states of the world that may occur.

In this manner, one has that $(\tilde{P}_m^{-1})_{m\in M\cup\{m_r\}}$ is potentially generalized Bayes plausible where $\tilde{P}_{m_r}^{-1} = (p^k_{m_r})_{k\geq 2}$. Using Lemma 1, there exists a set of bayesian devices $(\pi_k)_{k\geq 2}$ which leads to these posteriors. Let π'_1 be the bayesian device that sends message $m \in M$ with probability $\varepsilon\pi_1(m)$ and message m_r with probability $1-\varepsilon$. By construction, the value of $\Pi = (\pi_k)_{k\in K})$ would therefore be equal to $\varepsilon\overline{V}(p_0) + (1-\varepsilon)v(a_0, p_0)$.

In either case then, it is possible to create an ambiguous communication device such that its value, when computed with regard to π_1 is arbitrarily close to $\overline{V}(p_0)$.

Finally, using corollary 1, one can construct an ambiguous communication device whose (maxmin) value is arbitrarily close to $\overline{V}(p_0)$, which ends the proof.

Only if. In this section, we show that $\overline{V}(p_0)$ is the maximum value that can be obtained.

Let $\Pi = co(\pi_k)_K$ be an ambiguous device. Let $P_m = co(p^k_m)$ be the posterior sets resulting from this ambiguous communication device and τ^k the distribution over messages from bayesian device π_k. Let $v^* = \min_k \sum_m \tau^k(m)v_k(P_m)$ be the value of said device. Assume without loss of generality that $1 \in \underset{k}{\mathsf{argmax}}(\sum_m \tau^k(m)v_k(P_m))$.

$$\begin{aligned}
v^* &= \min_k \sum_m \tau^k(m)v_k(P_m) \\
&\leq \max_k \sum_m \tau^k(m)v_k(P_m) \\
&\leq \sum_m \tau^1(m)v_1(P_m) \\
&\leq \max_{P_m^{-1}} \sum_m \tau^1(m)v_1(co(p^1_m, P_m^{-1})) \\
&\leq \max_{P_m^{-1}} V_1(co(p_0, P_m^{-1})) \\
&\leq \overline{V}(p_0)
\end{aligned}$$

$\square$

A.5 Proof of Lemma 3

Proof. Consider an interval of length l and we have found above the optimal c that maximizes the expected payoff of Sender, given that Receiver optimally responds. Let us define

$$\begin{aligned}
\tilde{V}_1(l) &\equiv \int_0^l v\left(a_1(l), \omega\right) d\omega = -\frac{1}{12} l^3 - \left(b - \frac{l}{6}\right)^2 l. \\
\tilde{V}_2(l) &\equiv \int_0^l v\left(a_2(l), \omega\right) d\omega = -\frac{1}{12} l^3. \\
\tilde{V}(l) &\equiv \int_0^l v\left(a^*(l), \omega\right) d\omega.
\end{aligned}$$

Thus,

$$\tilde{V}(l) = \begin{cases} \tilde{V}_1(l), & \text{if } l \leq 6b; \\ \tilde{V}_2(l), & \text{if } l > 6b. \end{cases}$$

Thus, $\tilde{V}$ is the contribution to Sender's expected payoff from an interval of length l with c optimally chosen by Sender.

Note the function $\tilde{V}$ satisfies

$$\begin{aligned}
\tilde{V}_1'(l) &= -\frac{1}{3}(l-b)^2 - \frac{2}{3}b^2 \\
\tilde{V}_2'(l) &= -\frac{l^2}{4} \\
\tilde{V}_1''(l) &= \frac{2}{3}(b-l) \\
\tilde{V}_2''(l) &= -\frac{1}{2}l.
\end{aligned}$$

Furthermore,

$$\tilde{V}_1'(6b) = \tilde{V}_2'(6b)$$

Therefore, we conclude that the function $\tilde{V}'$ is (i) continuously differentiable, (ii) decreasing on $[0, b)$ and increasing on $(b, 1]$, and (iii) symmetric about b on $[0, 2b]$, i.e., $\tilde{V}'(l) = \tilde{V}'(2b-l)$ for all $l \in [0, 2b]$. The function $\tilde{V}$ is twice continuously differentiable, convex on $[0, b]$ and concave on $[b, 1]$.

For an arbitrary simple ambiguous communication device $(M, \Pi(\mathbf{y}, \mathbf{c}))$ with $2n$ messages,

$$V(\Pi(\mathbf{y}, \mathbf{c})) \leq V(\Pi(\mathbf{y}, \mathbf{c}^*(\mathbf{y}))) = \sum_{i=1}^{n} \tilde{V}(l_i)$$

We will discuss in two cases.

Case (i): If $l_i \geq b$ for all $i = 1, \ldots, b$, then by concavity of $\tilde{V}$ on $[b, 1]$

$$\sum_{i=1}^{n} \tilde{V}(l_i) \leq n\tilde{V}(\frac{1}{n}) = V(\Pi(1/n, c^*(n)))$$

Hence Sender would prefer the symmetric ambiguous communication device $\Pi(1/n, c^*(n))$.

Case (ii) If $l_i < b$ for some i. Since the order of the intervals $I_1, \ldots, I_n$ does not matter for V, without loss of generality we assume $l_1 \leq ... \leq l_i \leq b \leq l_{i+1} \leq \ldots \leq l_n$.

Start from I_1 and move right towards intervals with higher indices. If interval I_i has length $l_i \geq b$, then move to the next without change. If $l_i < b$, then make either of the two operations on intervals I_i and I_{i+1}: (i) If $l_i + l_{i+1} < 2b$ then combine I_i and I_{i+1} into one and relabel the new interval $[y_{i-1}, y_{i+1})$ as I'_i and continue from the new interval I'_i; (ii) If $l_i + l_{i+1} \geq 2b$, then adjust the intervals to $I'_i = [y_{i-1}, y_{i-1} + b)$ and $I'_{i+1} = [y_{i-1} + b, y_{i+1})$ and then move right to the interval I_{i+2} . Repeat the operations until the last interval I_n. And if $l_n < b$, either combine or adjust I_n with its left adjacent interval. This leads to new partition $\{I'_1, \ldots, I'_{n'}\}$ of $[0, 1]$ such that each cell have length no less than b, i.e., $l'_i \geq b$ for all $i = 1, \ldots, n'$. Then we are back to case (i).

We finish the proof by showing that either combining or adjusting two intervals as defined above increase sender's ex-ante utility V. Formally, for all $l < b$ and $l' \in [0, 1]$

$$\tilde{V}(l + l') \geq \tilde{V}(l) + \tilde{V}(l') \qquad \text{if } l + l' < 2b \tag{6}$$

and

$$\tilde{V}(b) + \tilde{V}(l + l' - b) \geq \tilde{V}(l) + \tilde{V}(l') \qquad \text{if } l + l' \geq 2b \tag{7}$$

To prove inequality (6), note that if $l + l' < 2b$, then

$$\tilde{V}(l + l') = \tilde{V}(l') + \int_{l'}^{l+l'} \tilde{V}'(s)ds$$

If $l' \geq b$, let $\delta = 2b - l - l' > 0$ and

$$\tilde{V}(l) = \int_{0}^{l} \tilde{V}'(s)ds = \int_{2b-l}^{2b} \tilde{V}'(s)ds = \int_{l'}^{l+l'} \tilde{V}'(s + \delta)ds < \int_{l'}^{l+l'} \tilde{V}'(s)ds$$

where the second equality follows from the property $\tilde{V}'(l) = \tilde{V}'(2b - l)$ for all $l \in [0, b]$, and the inequality follows from $\tilde{V}'$ is strictly decreasing on $(b, 1]$.

If $l' < b \leq l + l'$, then

$$\begin{aligned}
\int_{l'}^{l+l'} \tilde{V}'(s)ds &= \int_{l'}^{b} \tilde{V}'(s)ds + \int_{b}^{l+l'} \tilde{V}'(s)ds \\
&> \int_{l-b+l'}^{l} \tilde{V}'(s)ds + \int_{2b-(l-b+l')}^{2b} \tilde{V}'(s)ds \\
&= \int_{l-b+l'}^{l} \tilde{V}'(s)ds + \int_{0}^{l+l'+b} \tilde{V}'(s)ds = \int_{0}^{l} \tilde{V}'(s)ds = \tilde{V}(l)
\end{aligned}$$

where the inequality follows from V' is decreasing on $[b,1]$ and strictly increasing on $[0,b)$, and where the penultimate equality follows from the property $\tilde{V}'(l) = \tilde{V}'(2b-l)$ for all $l \in [0,b]$.

If $l + l' \leq b$, then

$$\int_{l'}^{l+l'} \tilde{V}'(s)ds > \int_0^l \tilde{V}'(s)ds = \tilde{V}(l)$$

To prove inequality (7), note that if $l + l' \geq 2b$,

$$\tilde{V}(b) - \tilde{V}(l) = \int_l^b \tilde{V}'(s)ds = \int_b^{b+(b-l)} \tilde{V}'(s)ds$$

where the last $=$ is due to the property $\tilde{V}'(l) = \tilde{V}'(2b-l)$ for all $l \in [0,b]$.

$$\tilde{V}(l') - \tilde{V}(l+l'-b) = \int_{(l+l'-b)}^{l'} \tilde{V}'(t)dt < \int_b^{b+(b-l)} \tilde{V}'(s)ds$$

where the inequality is due to $\tilde{V}'$ is decreasing on $[b,1]$.

□

A.6 Proof of Proposition 3

Proof. By symmetry, and following the calculation preceding Lemma 3, we have that Receiver's optimal action upon observing either message $m_{i,A}$ or message $m_{i,B}$ is

$$a_i^* = \begin{cases} \frac{i-1}{n} + \frac{1}{3n} + b, & \text{if } 1/n \leq 6b; \\ \frac{i-1}{n} + \frac{1}{2n}, & \text{if } 1/n \geq 6b. \end{cases}$$

Note that Sender's expected payoff can be written

$$V(\Pi(c), b) = -\sum_{i=1}^{n} \int_{(i-1)/n}^{i/n} [a_i^* - \omega]^2 \, d\omega.$$

By symmetry, when $b \leq 1/(6n)$, Sender's expected payoff from the optimal simple ambiguous communication device is

$$\begin{aligned} V(\Pi(c), b) &= -n \int_0^{1/n} \left[\frac{1}{2n} - \omega\right]^2 d\omega, \\ &= -\frac{1}{12n^2}, \\ &< -\frac{1}{36n^2} \leq -b^2. \end{aligned}$$

So it cannot dominate full disclosure of information without ambiguity. When $b \geq 1/(6n)$, Sender's expected payoff from the optimal simple ambiguous communication device is

$$\begin{aligned} V(\Pi(c), b) &= -n \int_0^{1/n} \left[\frac{1}{3n} + b - \omega \right]^2 d\omega, \\ &= -n \int_0^{1/n} \left[\frac{1}{2n} - \omega + b - \frac{1}{6n} \right]^2 d\omega, \\ &= -\left[b^2 - \frac{1}{3n} b + \frac{1}{9n^2} \right], \end{aligned}$$

which is greater than or equal to Sender's payoff under full disclosure if and only if

$$b \geq \frac{1}{3n}.$$

Now, we consider Sender's optimal choice of n, note that as long as $b \geq 1/(3n)$, or $n \geq 1/(3b)$, Sender's expected payoff is better than or equal to that under full disclosure, while if $b \leq 1/(3n)$, or $n \leq 1/(3b)$, his expected payoff is worse.

Therefore, to maximize his expected payoff, Sender would choose $n \geq 1/(3b) > 1/(6b)$, which implies that

$$a_1^* \equiv x_1^* + b = \frac{1}{3n} + b,$$

and Sender's expected payoff is

$$\begin{aligned} V(\Pi(c), b) &= -\left[b^2 - \frac{1}{3n} b + \frac{1}{9n^2} \right], \\ &= -\left(\frac{1}{3n} - \frac{b}{2} \right)^2 - \frac{3}{4} b^2, \end{aligned} \tag{8}$$

which is maximized when

$$\frac{1}{3n} = \frac{b}{2},$$

or

$$n = \hat{n}(b) \equiv \frac{2}{3b}.$$

So the optimal choice of n for Sender, $n^*(b)$, is the integer closest to $\hat{n}(b) \equiv 2/(3b)$. Note that Sender's highest expected payoff is

$$-\frac{3}{4} b^2$$

when $\hat{n}(b)$ is an integer and greater than or equal to

$$-\frac{13}{16} b^2,$$

when $\hat{n}(b) \geq 1$, and greater than or equal to

$$-b^2$$

even when $\hat{n}(b) < 1$ (or $b > 2/3$) because when $n = 1$

$$V(\Pi(c), b) \quad = \quad -\frac{1}{9} + \frac{1}{3}b - b^2 > -b^2.$$

Our conclusion is that using ambiguous messages definitely improves upon Sender's expected payoff under full disclosure, $-b^2$.

Now we want to check Receiver's participation constraint: at Sender optimal simple Ellsberg randomization device, Receiver's ex-ante utility is higher than the case when she gets no information at all. This condition guarantees Receiver still prefers receiving the ambiguous message, even if she could opt out ex-ante.

Let π_0 denote the null information. Receiver's optimal action is $\hat{a}(\pi_0) = \frac{1}{2} + b$. Her ex-ante utility is

$$U(\pi_0, b) = -\int_0^1 (\frac{1}{2} - \omega)^2 d\omega = \int_{\frac{1}{2}}^{-\frac{1}{2}} x^2 dx = -\frac{1}{12}$$

Consider the n-equal partition simple ambiguous communication device described above. Note that $b \geq 1/(6n)$, $c_1^* = 0$ and $x_1^* + b = \frac{1}{3n} + b$.

When $b \geq 1/(6n)$, the receiver's ex-ante expected payoff from the optimal ambiguous communication device is

$$\begin{aligned} U(\Pi(c), b) &= -n \int_0^{1/n} \left[\frac{1}{3n} - \omega\right]^2 d\omega, \\ &= -n \int_0^{1/n} \left[\frac{1}{9n^2} - \frac{2\omega}{3n} + \omega^2\right] d\omega, \\ &= -\frac{1}{9n^2}, \end{aligned}$$

which is greater than or equal to Receiver's payoff at no information if and only if $n \geq 2$.

Coupled with our conclusion that Sender's payoff can never exceed the full-disclosure payoff when $n \leq 1/(3b)$, if Receiver's participation constraint is to be respected, then Sender finds it optimal to fully disclose information to Receiver if and only if $n^*(b) = 1$, which, by (8), holds when

$$\frac{1}{3 \cdot 1} - \frac{1}{3\hat{n}(b)} \leq \frac{1}{3\hat{n}(b)} - \frac{1}{3 \cdot 2},$$

or

$$b \geq \frac{1}{2}.$$

When $n^*(b) \geq 2$, Sender will choose the simple ambiguous communication device given above. $\square$

A.7 Proof of Proposition 3

(Dynamic Consistency $\Leftarrow$ Rectangularity) in Epstein and Schneider (2003).

(Dynamic Consistency $\Rightarrow$ Rectangularity) By contradiction, if Π is not rectangular then P_0 the set of priors defined over $\Omega \times M$ is not rectangular à la Epstein and Schneider (2003). Thus, from Epstein and Schneider (2003), there exists f_1 and f_2 two acts in $\mathbb{R}^{\Omega \times M}$ such that $f_1 \succ f_2$ and for all message m, $f_1 \prec_{\Omega \times \{m\}} f_2$. Let $a_{k,m}$ be an action that generates utilities $u(a_{k,m}, \omega) = f_k(\{\omega\} \times \{m\})$. Thus, whereas the optimal ex post strategy is given by $(a_{2,m})_m$, it is beaten by $(a_{1,m})_m$ at the ex ante stage.

A.8 Proof of Proposition 5

Proof. Suppose that $R = \bar{R}$ and $(Q^*_m)_{m \in M}$ is potentially generalized Bayes plausible and hence satisfies conditions in Lemma 1. For any $q^*_m \in Q^*_m$, there exists $(q^*_m)_{m \in M} \in (Q^*_m)_{m \in M} \subseteq (co(P_m))_{m \in M}$ and $\bar{\tau} \in \Delta M$ such that $\sum_m \bar{\tau}(m) q^*_m = p_0$. By definition, $(\bar{\tau}, \mathbf{q}^*) \in co(\bar{R})$.

Then Sender can do at least as well with Bayesian persuasion $(\bar{m}, \mathbf{q}^*)$ compared to ambiguous persuasion with signal R. To see this,

$$\begin{aligned}
v(\bar{\tau}, \mathbf{q}^*) &= \sum_m \bar{\tau}(m) E_{q^*_m} v(\hat{a}(q^*_m), \omega) \\
&= \sum_m \bar{\tau}(m) E_{q^*_m} v(\hat{a}(P_m), \omega) \\
&\geq \min_{(\tau, \mathbf{p}) \in co(\bar{R})} \sum_m \tau(m) E_{p_m} v(\hat{a}(P_m), \omega) \\
&= \min_{(\tau, \mathbf{p}) \in \bar{R}} \sum_m \tau(m) E_{p_m} v(\hat{a}(P_m), \omega) \\
&= \min_{(\tau, \mathbf{p}) \in R} \sum_m \tau(m) E_{p_m} v(\hat{a}(P_m), \omega) \\
&= v(R). \quad \square
\end{aligned}$$

A.9 Proof for the example of section 6.3

Proposition 2. *Assume $\Pi = conv(\underline{\pi}, \overline{\pi})$ is a device using only three messages for two states of the world. probabilities are given by the probability of the high state occurring. A low message yields the unique posterior $\mathbf{p}_l$, the middle message yields the posterior set $[\underline{p}, \overline{p}]$ and the high message yields the unique posterior p_h with $p_l < \underline{p} < \overline{p} < p_0$. Let $\underline{\pi}$ be the probabilistic device that yields the posterior $\underline{p}$ at the middle message. Then Π is valuable if and only if:*

- $p_0 < \overline{p} \Rightarrow \frac{p_h - \overline{p}}{\overline{p} - \underline{p}} \geq \frac{\underline{\pi}(m)}{\underline{\pi}(h)}$ *and*

- $p_0 > \underline{p} \Rightarrow \frac{\underline{p} - p_l}{\overline{p} - \underline{p}} \geq \frac{\overline{\pi}(m)}{\overline{\pi}(l)}$

The following provides an intuition of this result. Consider for example Figure 6 which shows the utilities from two actions a_0 and a_1, assuming a_0 is picked at p_0. In this case, a_1 is picked both at message "m", the ambiguous outcome, and at message "h", the risky outcome. G denotes the gain at message "h" from playing a_1 instead of a_0. Likewise, L is the loss of picking a_1 instead of a_0 at message "m" when evaluated at $\underline{p}$. For the net effect to be positive, the weights attributed to each outcome ex ante, which can be shown to be $\underline{\pi}(m)$ and $\underline{\pi}(h)$, must be so that $\underline{\pi}(h)G - \underline{\pi}(m)L \geq 0$ or in other terms that the ratio of gains to losses G/L must be greater than the relative frequencies of losing vs winning $\underline{\pi}(m)/\underline{\pi}(h)$. From the figure and Thales' theorem, we have that $G/L = G'/L'$ which yields the left hand side of the inequality in the proposition as one can further prove that this is the worst possible case given Π. Thus it is possible to have a valuable non-rectangular signal by making sure there is some non-ambiguous information that is sufficiently better than the present ambiguous information (sufficiently more spread out) that is attained sufficiently often.

Proof. I first show that the result holds when there are only two actions a_0 and a_1 available and then extend the result to more actions.

I assume here that Π is embedded and that $p_0 \leq \overline{p}$. I then compute the value of information dependent on the utilities and show that this is positive if and only if $\frac{p_h - \underline{p}}{\overline{p} - \underline{p}} \geq \frac{\underline{\pi}(m)}{\underline{\pi}(h)}$.

Without loss of generality, let a_0 yield the utilities $a > 0$ in ω_2 and 0 otherwise. Let a_1 yield the utilities b in ω_2 and c in ω_1 such that $c \geq 0$ and $b \leq a$. If this were not the case, then one action would be dominated by the other and the same action would be played.

Assume now that $p = \frac{a-b}{a+c-b}$, the probability under which both actions are indifferent to the decision maker, is not in $[\underline{p}; \overline{p}]$. In this case, ambiguity has no bearing and so the value of information is positive.

Assume now that $p \in [\underline{p}; \overline{p}]$. Assume further that $p \geq p_0$. At equilibrium, the actions chosen by the decision maker are a_0 without information and at message "l", a_1 at message "h" while he plays a_1 with probability $\sigma(a, b, c) = \frac{a}{a+c-b}$ if $b \leq c$ and 1 otherwise.

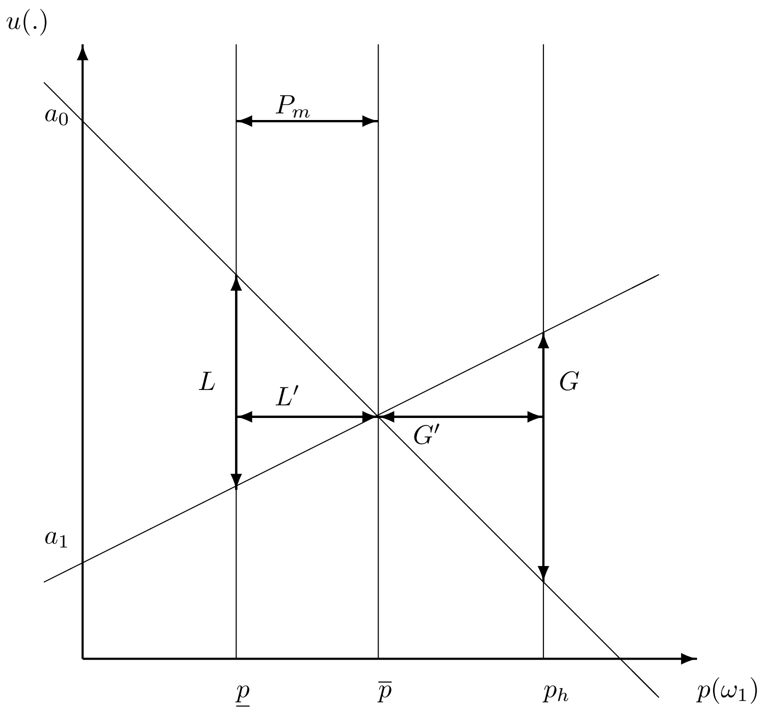

Figure 6: Gains and Losses from Embedded Information

Let $G(a,b,c)$ be the decision maker's gain over playing a_0 at message "h". Then:

$$G(a,b,c) = b - a + p_h(a + c - b)$$

Given $p_h \geq p = \frac{a-b}{a+c-b}$ by assumption, this is indeed a gain. Let $L(a,b,c)$ be the loss incurred at "m" when evaluated at $\tilde{p}$:

$$L(a,b,c,\tilde{p}) = (1-\tilde{p})a - \sigma(a,b,c).[\tilde{p}c + (1-\tilde{p})b] - (1-\sigma(a,b,c))(1-\tilde{p})a$$

I now show that $U_0 > U(\Pi)$ is equivalent to $\underline{\pi}(h)G(a,b,c) - \underline{\pi}(m)L(a,b,c,\underline{p}) < 0$. Compute the ex ante utility:

$$U(\Pi) = \min_{\pi \in \Pi} \pi(l)u(a_0, p_l) + \pi(h)u(a_1, p_h) + \pi(m)u(\sigma, p_\pi)$$

First assume that $U(\Pi)$ is minimized at $\overline{\pi}$:

$$U(\overline{\pi}) = \overline{\pi}(l)u(a_0, p_l) + \overline{\pi}(h)\underbrace{u(a_1, p_h)}_{>u(a_0,p_h)} + \overline{\pi}(m)\underbrace{u(\sigma, \overline{p})}_{\geq u(a_0,\overline{p})}$$

Thus, this would imply $U(\Pi) > U_0$. Assume now that π is such that $p_\pi \neq \underline{p}$. Given $\Pi = conv(\underline{\pi}, \overline{\pi})$, there exists λ such that $\pi = \lambda\underline{\pi} + (1-\lambda)\overline{\pi}$. Thus by linearity:

$$U(\lambda) = \lambda U(\underline{\pi}) + (1-\lambda)U(\overline{\pi})$$

Thus given $U(\overline{\pi}) > 0$, $U(\lambda) < U_0 \Rightarrow U(\underline{\pi}) < U_0$. Computing $U(\underline{\pi}) - U_0$ yields $\underline{\pi}(h)G(a,b) - \underline{\pi}(m)L(a,b,c,\underline{p})$. I now shorten $L(a,b,c,\underline{p})$ to $L(a,b,c)$ for brevity.

G is increasing in c as $c > 0$. If $b \geq c$, then $\sigma(a,b,c) = 1$ and L is decreasing in c. If $c \geq b$, then[23]:

$$L(a,b,c) = \sigma(a,b,c)[a - b - \underline{p}(a + c - b)]$$

Thus, given $\sigma(a,b,c)$ is decreasing in c, L is the product of two decreasing functions in c so is decreasing in c. Thus, one has:

$$\underline{\pi}(h)G(a,b,c) - \underline{\pi}(m)L(a,b,c) < 0 \Rightarrow \underline{\pi}(h)G(a,b,\frac{(1-\overline{p})(a-b)}{\overline{p}}) - \underline{\pi}(m)L(a,b,\frac{(1-\overline{p})(a-b)}{\overline{p}}) < 0$$

As $\frac{(1-\overline{p})(a-b)}{\overline{p}}$ is the smallest value of c compatible with $p \in [\underline{p};\overline{p}]$. Note that in this case, $\sigma(a,b,c) = 1$. This gives us new functions:

$$G(a,b) = b - a + p_h(a + \frac{(1-\overline{p})(a-b)}{\overline{p}} - b) = (a-b)\left[\frac{p_h}{\overline{p}} - 1\right]$$

23
$$\begin{aligned} L(a,b,c) &= (1-\underline{p})a - \tfrac{a}{a+c-b}.[\underline{p}c + (1-\underline{p})b] - (1 - \tfrac{a}{a+c-b})(1-\underline{p})a \\ &= \tfrac{a}{a+c-b}[(1-\underline{p})(a+c-b) - \underline{p}c - (1-\underline{p})b - (1-\underline{p})(a+c-b) + (1-\underline{p})a] \\ &= \sigma(a,b,c)[-\underline{p}c - (1-\underline{p})b + (1-\underline{p})a] \\ &= \sigma(a,b,c)[a - b - \underline{p}(a+c-b)] \end{aligned}$$

and:

$$L(a,b) = a - b - \underline{p}(a + \frac{(1-\overline{p})(a-b)}{\overline{p}} - b) = (a-b)\left[1 - \frac{\underline{p}}{\overline{p}}\right]$$

Given $\underline{\pi}(h)G(a,b) - \underline{\pi}(m)L(a,b) < 0 \Leftrightarrow \frac{G(a,b)}{L(a,b)} < \frac{\underline{\pi}(m)}{\overline{\pi}(h)}$, one therefore has that there exists utilities with negative value of information for Π if and only if:

$$\frac{G(a,b)}{L(a,b)} = \frac{\left[\frac{p_h}{\overline{p}} - 1\right]}{1 - \frac{\underline{p}}{\overline{p}}} = \frac{p_h - \overline{p}}{\overline{p} - \underline{p}} < \frac{\underline{\pi}(m)}{\overline{\pi}(h)}$$

Note if p had been smaller than p_0, the condition would be that $\frac{\overline{p}-p_l}{\overline{p}-\underline{p}} \geq \frac{\overline{\pi}(m)}{\overline{\pi}(l)}$.

In conclusion:

- If $p_0 < \overline{p}$ and $\frac{p_h - \overline{p}}{\overline{p}-\underline{p}} < \frac{\underline{\pi}(m)}{\overline{\pi}(h)}$, then one can construct utilities that yield negative value of information[24].
- Thus Π valuable implies condition 1) in the proposition. The exact same argument is used to show that Π valuable implies condition 2)
- If Conditions 1 and 2 apply, then as shown above, no utilities can yield negative information which proves the other direction of the proposition
- This has been proved when only two actions were available. To extend the result, realize that given d actions, one can construct similar decision problems for actions that yield the same payoffs as any mixed strategy among the d actions. Thus, if one were to take a_0 the action that yields the same payoffs as the action taken without information in the full game and a_1 the action that yields the payoffs of the mixed strategy taken in the original game at "m", then our result applies to this new game. In both games U_0 is left unchanged and $U(\Pi)$ is smaller in the small game as it has in effect stopped the decision maker from choosing optimally at "h" and "l" without modifying the payoffs at "m". Given the new $U(\Pi)$ is greater than U_0, then so must be the former.

□

[24]pick a and b randomly and $c = \frac{(1-\overline{p})(a-b)}{\overline{p}}$)

A.10 Proof of Proposition 1

Proof. For the first component, if $\hat{a}(P_m) = a_0$ then P_m is obviously value-increasing. Otherwise, assume by contradiction that there exists $p_m \in P_m \cap p^{-1}(a_0)$, then by definition of $p^{-1}(a_0)$, $u(a_0, p_m) > u(\hat{a}(P_m) \neq a_0, p_m)$.

For the second component. Let Π use only value-increasing messages. Then the value to the receiver is given by:

$$
\begin{aligned}
U(\Pi) &= \min_k \sum_m \tau_m^k \tilde{u}_k(P_m) \\
&= \min_k \sum_m \tau_m^k u(\sigma(P_m), p_m^k) \\
&\geq \min_k \sum_m \tau_m^k u(a_0, p_m^k) \\
&\geq \min_k \; u(a_0, p_0) \geq U_0
\end{aligned}
$$

□

References

Abdellaoui, M., A. Baillon, L. Placido, and P. P. Wakker (2011). The rich domain of uncertainty: Source functions and their experimental implementation. *American Economic Review 101*(2), 695–723.

Alonso, R. and O. Câmara (2016). Bayesian persuasion with heterogeneous priors. *Journal of Economic Theory 165*, 672 – 706.

Anscombe, F. J. and R. J. Aumann (1963). A definition of subjective probability. *The Annals of Mathematical Statistics 34*(1), pp. 199–205.

Aumann, R. and M. Maschler (1995). *Repeated Games with Incomplete Information.* The MIT Press.

Ayouni, M. and F. Koessler (2015). Hard evidence and ambiguity aversion. Working Paper.

Bade, S. (2011). Ambiguous act equilibria. *Games and Economic Behavior 71*, 246–260.

Blackwell, T. (2014, 14 September). The new drug wars: Brand-name pharma giants attack generic firms as competition grows. *National Post*.

Blume, A. and O. Board (2014). Intentional vagueness. *Erkenntnis 79*(4), 855–899.

Blume, A., O. Board, and K. Kawamura (2007). Noisy talk. *Theoretical Economics 2*(4), 395–440.

Bose, S., E. Ozdenoren, and A. Pape (2006, December). Optimal auctions with ambiguity. *Theoretical Economics 1*(4), 411–438.

Bose, S. and L. Renou (2014). Mechanism design with ambiguous communication devices. *Econometrica 82*(5), 1853–1872.

Bossaerts, P., P. Ghirardato, S. Guarnaschelli, and W. R. Zame (2010). Ambiguity in asset markets: Theory and experiment. *Review of Financial Studies 23*(4), 1325–1359.

Camerer, C. and M. Weber (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty 5*(4), 325–70.

Chow, C. C. and R. K. Sarin (2001). Comparative ignorance and the Ellsberg paradox. *Journal of Risk and Uncertainty 22*(2), 129–139.

Crawford, V. and J. Sobel (1982). Strategic information transmission. *Econometrica 50*(6), 1431–51.

Crémer, J., L. Garicano, and A. Prat (2007). Language and the theory of the firm. *The Quarterly Journal of Economics 122*(1), 373–407.

Di Tillio, A., N. Kos, and M. Messner (2015). The design of ambiguous mechanisms. Working Paper.

Epstein, L. G. and M. Schneider (2003). Recursive multiple-priors. *Journal of Economic Theory 113*(1), 1–31.

Epstein, L. G. and M. Schneider (2007). Learning under ambiguity. *Review of Economic Studies 74*(4), 1275–1303.

Epstein, L. G. and M. Schneider (2010). Ambiguity and asset markets. *Annual Review of Financial Economics 2*(1), 315–346.

Epstein, L. G. and K. Seo (2010, Sep). Symmetry of evidence without evidence of symmetry. *Theoretical Economics 5*(3), 313–368.

Fox, C. R. and A. Tversky (1995). Ambiguity aversion and comparative ignorance. *The Quarterly Journal of Economics 110*(3), 585–603.

Frankel, A. (2014, January). Aligned delegation. *American Economic Review 104*(1), 66–83.

Gajdos, T., T. Hayashi, J.-M. Tallon, and J.-C. Vergnaud (2008). Attitude toward imprecise information. *Journal of Economic Theory 140*(1), 27–65.

Gilboa, I. and M. Marinacci (2013). Ambiguity and the bayesian paradigm. In *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress of the Econometric Society*.

Gilboa, I. and D. Schmeidler (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics 18*(2), 141–153.

Halevy, Y. (2007). Ellsberg revisited: An experimental study. *Econometrica 75*(2), 503–536.

Ivanov, M. (2010). Informational control and organizational design. *Journal of Economic Theory 145*(2), 721 – 751.

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

Kellner, C. and M. T. Le Quement (2015). Modes of ambiguous communication. B*onn working paper*.

Kellner, C. and M. T. Le Quement (2016). Endogenous ambiguity in cheap talk. B*onn working paper*.

Klibanoff, P., M. Marinacci, and S. Mukerji (2005). A smooth model of decision making under ambiguity. *Econometrica 73*(6), 1849–1892.

Laclau, M. and L. Renou (2016). Public persuasion. Working paper.

Lipman, B. L. (2009). Why is language vague? *Unpublished paper, Boston University*.

Lipnowski, E. and L. Mathevet (2015). Disclosure to a psychological audience. NYU Working Paper.

Lopomo, G., L. Rigotti, and C. Shannon (2011). Knightian uncertainty and moral hazard. *Journal of Economic Theory 146*(3), 1148 – 1172. Incompleteness and Uncertainty in Economics.

Lopomo, G., L. Rigotti, and C. Shannon (2014). Uncertainty in mechanism design. Working Paper.

Mukerji, S. and J.-M. Tallon (2004). An overview of economic applications of David Schmeidler's models of decision making under uncertainty. In *Uncertainty in Economic Theory,* (I. Gilboa, ed.), New York: Routledge.

Pinker, S. (2015). *The Sense of Style: The Thinking Person's Guide to Writing in the 21st Century!* Viking Press–Penguin Books.

Pires, C. P. (2002). A rule for updating ambiguous beliefs. *Theory and Decision 53*(2), 137–152.

Riedel, F. and L. Sass (2014). Ellsberg games. *Theory and Decision 76*(4), 469–509.

Schlee, E. E. (1997). The sure thing principle and the value of information. *Theory and decision 42*(1), 21–36.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica 57*(3), 571–587.

Seidenfeld, T. and L. Wasserman (1993). Dilation for sets of probabilities. *The Annals of Statistics 21*(3), 1139–1154.

Siniscalchi, M. (2011). Dynamic choice under ambiguity. *Theoretical Economics 6*(3), 379–421.

Sobel, J. (2013). Giving and receiving advice. In D. Acemoglu, M. Arellano, and E. Dekel (Eds.), *Advances in Economics and Econometrics: Tenth World Congress*, pp. 305–341. Cambridge: Cambridge University Press.

Sobel, J. (2015). Broad terms and organizational codes. Technical report, University of California, San Diego.

Wolitzky, A. (2016). Mechanism design with maxmin agents: Theory and an application to bilateral trade. *Theoretical Economics 11*(3), 971–1004.