Accepted Manuscript

Ambiguous persuasion

Dorian Beauchêne, Jian Li, Ming Li

PII: S0022-0531(18)30665-3
DOI: https://doi.org/10.1016/j.jet.2018.10.008
Reference: YJETH 4837

To appear in: *Journal of Economic Theory*

Received date: 26 January 2017
Revised date: 11 October 2018
Accepted date: 21 October 2018

Please cite this article in press as: Beauchêne, D., et al. Ambiguous persuasion. *J. Econ. Theory* (2018), https://doi.org/10.1016/j.jet.2018.10.008

Ambiguous Persuasion*

Dorian Beauchêne[†] Jian Li[‡] Ming Li[§]

October 10, 2018

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, a Sender may choose to use ambiguous communication devices. Our main result characterizes the value of optimal ambiguous persuasion, which is often higher than what is feasible under Bayesian persuasion. We characterize posteriors that are potentially plausible when they are generated by 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 standard Bayesian persuasion. We also show that the use of synonyms are a necessary

*We are grateful to the Editor, Marciano Siniscalchi, and two anonymous referees for very constructive comments and suggestions that have significantly improved the paper. We are also grateful to Pierre Fleckinger, Françoise Forges, Sidartha Gordon, Yingni Guo, Sean Horan, Rafael Hortala-Vallve, Maxim Ivanov, Frédéric Koessler, Marie Laclau, Laura Lasio, Elliot Lipnowski, Sujoy Mukerji, Peter Norman, Eduardo Perez-Richet, Ludovic Renou, Joel Sobel, Jean-Marc Tallon, Vassili Vergopoulos, and two anonymous referees for their very helpful comments. We also thank 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, RES-York Symposium on Game Theory 2017, D-TEA 2018, BEAT 2018, Carleton University, Kyoto University, McMaster University, McGill University, UC Berkeley, National University of Singapore, Université de Montreal, and Université Laval 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), "Communication avec un public averse à l'ambiguïté" (FRQSC 2018-NP-206978, Jian Li), and "Scientific research, conflicts of interest, and disclosure policy" (SSHRC Insight Development 430-2016-00444, Ming Li). Dorian Beauchêne conducted part of the research while writing his Ph.D. dissertation at Paris School of Economics, and Jian Li conducted part of the research at McGill University. In addition, Beauchêne thanks CIREQ and Ming Li thanks ISER at Osaka University for their hospitality, where they have conducted part of the research.

[†]MAPP Economics. E-mail: dorianbeauchene@gmail.com

[‡]Corresponding author. School of Economics, Shanghai University of Finance and Economics. E-mail: jianli02@gmail.com.

[§]Concordia University and CIREQ. Email: ming.li@concordia.ca.

property of optimal and beneficial ambiguous persuasion. We consider two applications, including the well-known uniform-quadratic example. Our analysis provides a justification for how ambiguity may emerge endogenously in persuasion.

JEL Classification Numbers: C72, D81, D83.

KEYWORDS: Bayesian persuasion, ambiguity averision, information transmission.

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. 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 (him)self 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, whom we will refer to as "Sender" and "Receiver" hereafter. 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

[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.

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.

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 generic 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 her 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 (EU) 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 via the full Bayesian rule (à la Pires 2002; 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 utilities.

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 by examining two aspects of the environment. First, we present a "*splitting lemma*," namely, a characterization of the possible profiles of posterior sets that are achievable with ambiguous devices, which precisely pins down the extent to which ambiguous devices expand beyond the Bayes plausibility condition given by Kamenica and Gentzkow

(2011). Second, we demonstrate that, when using an ambiguous device consisting of a set of probabilistic devices with different expected payoffs, Sender can approximately achieve the highest payoff among them. He does so through mixing probabilistic devices via the use of *synonyms*, which are messages that induce the same set of posteriors and therefore lead to the same action from Receiver. In effect, Sender can always hedge against ambiguity created by his own choices. Consequently, with the construction of synonyms, in optimal ambiguous persuasion, every probabilistic device will give Sender the same expected payoff. Building on these two intermediate findings, our main result, Proposition 1, characterizes Sender's optimal payoff as the *maximal* projection of the *concave closure* of an interim value function, which is defined as Sender's interim expected payoff according to one particular posterior and given that this posterior belongs to Receiver's posterior set. Therefore, our interim value function depends on a vector/set of posteriors, which are generated by full Bayesian updating based on the ambiguous device. Our result is reminiscent of the "concavification" result of Kamenica and Gentzkow (2011), but the upper bound of the projected concave closure, as a consequence of synonyms and hedging, is somewhat surprising, which brings fresh insights that are quite apart from Kamenica and Gentzkow's.

Further, we show that the use of synonyms (albeit in a weaker sense) is also necessary for optimal ambiguous persuasion. We demonstrate that generically either the optimal ambiguous devices use weak synonyms, which are messages that elicit the same Receiver action, or ambiguous persuasion is not more valuable than Bayesian persuasion.

We then proceed to explore the structure of optimal ambiguous device 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 it has the highest worst-case payoff for Receiver. Even though this action might not be ex ante optimal, our splitting lemma suggests that, through designing an optimal ambiguous device, Sender can create maximum (ex post) ambiguity at all messages and succeed in persuading Receiver to always take the safe action, which is not possible under standard Bayesian persuasion.

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, a recent paper by Laclau and Renou (2016) is related to our work. Laclau and Renou (2016) consider *public persuasion*, where Sender sends the same (probabilistic) 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 Gentzkow's (2011) concept of Bayes plausibility, and provide a version of "concavification" for the case of multiple priors and a single probabilisic communication device. We differ from their approach by focusing on the case when there is no prior ambiguity while Sender may use an ambiguous communication device. As a result, hedging (via the use of synonyms) plays a unique role in our characterization. We also assume both Sender and Receiver can be ambiguity averse to ensure symmetric information.

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 providing 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 (2017) solve a simplified two actions/two states game where ambiguity is present in Receiver's priors but not as a strategic choice of Sender. They show that Sender would use more messages under this assumption than with the regular Bayesian prior. In a follow-up paper, Kellner and Le Quement (2018) 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 communication between Sender and Receiver. These two papers' approach differ from ours in that they do not assume that 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]

[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. Relatedly, Lipnowski and Mathevet (2017) consider a model of Bayesian persuasion where Receiver has psychological preferences, where beliefs directly enter Receiver's payoff function.

[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

The introduction of ambiguity aversion into game theoretical models are supported by recent experimental findings that people dislike betting on an event with unknown probability.[6] This phenomenon has attracted significant interest in theory and applications.[7] The literature that explores the role of ambiguity aversion in games and mechanisms has been steadily growing (Ayouni and Koessler, 2017; Bose et al., 2006; Di Tillio et al., 2016; Frankel, 2014; Lopomo et al., 2011, 2014; Wolitzky, 2016). 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. Their key insight is that the designer can cleverly use these devices to exploit the ambiguity aversion of the agents. They show 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 Sender. Section 3 introduces the framework and presents the persuasion game. Section 4 characterizes the value of an optimal ambiguous communication device, which is the maximal projection of the concave closure of sender's interim expected utility function, and demonstrates how to construct an optimal ambiguous device. Section 5 provides two examples in which ambiguous communication devices 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 persuasion. 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 a physician. The physician could choose between two actions: prescribing the brand name producer's drug or prescribing a generic competitor of it. 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 the common prior, the brand name drug is less risky than the generic drug in the following sense: the brand name drug

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).

[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.

is always effective and produces some deterministic treatment utility u_H to the physician; the generic drug's effectiveness is uncertain: in one state denoted as $\omega =$ "Effective," it is as effective as the brand name drug and produces the same high utility u_H; in the other state, denoted as $\omega =$ "Ineffective," it is of low quality and produces a low utility u_L ($< u_H$) (due to unintended side effect, lower effectiveness).[8] Normalize and assume the unit price of the generic drug is 0 and that of the brand name drug is $c > 0$. Assume that Receiver, the physician, is well-intentioned and so her payoff is the treatment utility from the drug minus its cost; while Sender, the brand name drug producer, always prefers that its own drug be 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 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 $p(\text{Inffective}) = p_0 \in (0,1)$. Suppose $c/p_0 > u_H - u_L > c > 0$, 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) Propositions 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

$$\begin{aligned} \pi(e|\text{ Effective}) &= \frac{p^* - p_0}{p^*(1-p_0)}, \qquad \pi(i|\text{ Effective}) = \frac{p_0(1-p^*)}{p^*(1-p_0)}; \\ \pi(e|\text{ Ineffective}) &= 0, \qquad \pi(i|\text{ Inffective}) = 1, \end{aligned}$$

which induces a Bayes plausible distribution over posteriors

$$\begin{aligned} p_e(\text{Ineffective}) &= 0, \\ p_i(\text{Ineffective}) &= p^*, \end{aligned}$$

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

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

[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, $p_0 < p^*$.

Note that

$$Prob(m = i) = \frac{p_0}{p^*} < 1.$$

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 structure communication so that the physician is presented with multiple ways of interpretations of all messages, with the requirement that each way of interpretation is correct on average across messages. For example, the brand name producer could sponsor an experimental test of the generic drug and report the data, while remaining vague on how the data are generated and therefore how they should be interpreted. A formalization of this idea is a system with $M = \{e, i\}$ that admits two likelihood distributions,[10] $\Pi = \{\pi, \pi'\}$, where

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

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

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

As we will verify later, these posteriors are also "Bayes plausible." The main takeaway is that the drug company can design ambiguous messages so that choosing generic always looks extremely uncertain. Consequently, the MEU physician will prescribe the brand name drug regardless of the message observed. 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 now turn to setting up a general framework that formalizes and extends the discussion in the illustrating example above. We consider a persuasion game between Sender and Receiver ("he" and "she," respectively, as previously mentioned). Let Ω denote the set of states of the world and A the set of feasible actions. Assume Ω and A are compact subsets of the

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

Euclidean space. Receiver and Sender have continuous utility functions, denoted by u and v, respectively. Thus, $u(a,\omega)$ and $v(a,\omega)$ are respectively the utilities of Receiver and Sender, when Receiver's action is $a \in A$ and the state of the world is $\omega \in \Omega$.[11] Sender and Receiver share a common prior $p_0 \in \Delta\Omega$ with full support.[12]

Sender sends a message from a finite set M to Receiver, who then takes an action. Sender does so by committing to a communication device. A *probabilistic*/Bayesian communication device π is a function from states of the world to probability distributions over messages. Thus, $\pi(\cdot|\omega) \in \Delta M$ and $\pi(m|\omega)$ denotes the probability with which message m is sent in state ω. Further, let $\tau(m) = \sum_\omega p_0(\omega)\pi(m|\omega)$ denote the overall probability with which message m is sent. An *ambiguous* communication device consists of a finite set of probabilistic devices, denoted $(\pi_k)_K$, that are indexed by $k \in K = \{1, ..., |K|\}$.[13] Assume that the K devices share a common support on M, that is, for all k and $j \in K$ and $m \in M$, $\tau_k(m) > 0 \Rightarrow \tau_j(m) > 0$. Finally, denote by Π the convex hull of $(\pi_k)_K$,[14] which can be written as

$$\Pi = co((\pi_k)_K) = \left\{ \pi \in (\Delta M)^\Omega \;\middle|\; \pi = \sum_{k\in K} \lambda(k)\pi_k \text{ for some } \lambda \in \Delta K \right\}.$$

Our interpretation of Sender's commitment to an ambiguous communication device as follows. Sender designs K probabilistic devices $(M, (\pi_k)_K)$ and sends them to a credible third party. The third party first draws a ball from an Ellsberg urn containing a large number of balls labelled $1, 2, \ldots, K$. The label of the ball drawn determines which probabilistic device will be used. Then, the designated probabilistic device generates a message, which is observed by both Sender and Receiver. Yet, both players are ignorant about the label of the ball drawn and the composition of the labels of the balls in the Ellsberg urn.

We follow Pires (2002) and Epstein and Schneider (2007) and assume that both Sender and Receiver form their posteriors using the *full Bayesian updating rule*. That is, when the prior is p_0 and Sender chooses some ambiguous communication device (M, Π), upon receivering a message m Sender and Receiver update their beliefs probability by probability and form the following set of posteriors:

$$P_m = \{p_m^{\pi_k} \in \Delta(\Omega) : p_m^{\pi_k}(\cdot) = \frac{p_0(\cdot)\pi_k(m|\cdot)}{\int_{\omega'} \pi_k(m'|\omega')p_0(\omega')d\omega'}, \quad k \in K\}. \tag{1}$$

[11] If A or Ω is finite, endow it with the discrete topology.

[12] 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 ΔX to denote the set of probabilities on X endowed with the topology of weak convergence. The space of closed and convex subsets of ΔX 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.

[13] For simplicity of notation, we will use K to denote both the set of probabilistic devices and its cardinality, whenever there is no confusion.

[14] Considering of the convex hull of the probabilities is by convention of the MEU model, as only the convex hull of beliefs can be identified. In our model, equilibrium is unaffected by the choice between $(\pi_k)_K$ and Π, as only the extreme points of the set will be minimizing probabilities for an MEU agent.

Here $p_m^{\pi_k}$ denotes the Bayesian posterior associated with message m that is induced by probabilisitic device π_k.

The assumption that all probabilistic devices must have common support on M is a consequence of this full Bayesian updating rule (1). To see this, consider the induced set of ex-ante probabilities that belong to $\Delta(M \times \Omega)$, where

$$P = co\{p_k : p_k(m, \omega) = p_0(\omega)\pi_k(m|\omega), \quad k \in K\}.$$

For every message m, Bayesian updating is only defined if $p(\{m\} \times \Omega) > 0$. Hence full Bayesian updating requires that every message m either has zero probability for all $p \in P$ and thus is excluded from updating or has positive probability for all $p \in P$ and thus is updated probability by probability.

Another consequence of full Bayesian updating is that, when the communication device is ambiguous, the set of posteriors may become less precise than the prior *at all realized messages*. This is the case in our introductory example. As illustrated by the following example (Seidenfeld and Wasserman, 1993), it 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 1/2, that is, $p_0(H_2) = 1/2$. However, the players do not know how the two coin tosses are related (for instance, the coin might or might not have been simply turned over 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 $[0, 1/2]$. 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 set of posteriors is $P_{H_1}(H_2) = P_{T_1}(H_2) = [0, 1]$. Note the players' posterior beliefs about the event H_2 are dilated away from the prior for either realization of the first coin toss.

The rest of our model setup is similar to that of Kamenica and Gentzkow (2011). Our persuasion game consists of two stages. In the first/ex ante stage, Sender commits to an ambiguous communication device (M, Π). In the second/interim stage, both players observe the realized message and Receiver takes an action a. At the end of the game, the true state is revealed and payoffs are realized. We solve the game 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 maxmin expected utility; If there is a tie, Receiver chooses the action most preferred by Sender. (ii) In the ex ante stage, Sender chooses an ambiguous communication device (M, Π) that maximizes his ex ante maxmin expected utility.

In the interim stage, with multiple posteriors upon observing message m, Receiver's decision criterion is her interim maxmin expected utility with the probability-by-probability updated set of posteriors P_m:

$$U(a, P_m) = \min_{p_m \in co(P_m)} \mathbb{E}_{p_m}[u(a,\omega)] = \min_{p_m \in P_m} \mathbb{E}_{p_m}[u(a,\omega)].^{15}$$

Her optimal action is given by $a^*(P_m) \in \arg\max_{a\in A} U(a, P_m)$; when there are multiple maximizers, without loss assume Receiver picks the action preferred by Sender according to his expected utility computed by the first posterior[16], and this Sender-preferred optimal action is denoted by $\hat{a}(P_m)$.

It is 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 message. The property is called *language invariance* by Alonso and Câmara (2016), that is, for any two communication devices Π and Π', and any messages m and m', Receiver will always choose the same optimal action whenver $P_m = P_{m'}$.

In the ex-ante 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[17]

$$\sup_{(M,\Pi)} \min_{\pi\in\Pi} \mathbb{E}_{p_0}[\mathbb{E}_\pi[v(\hat{a}(P_m),\omega)|\omega]]. \tag{2}$$

We use V_0 to denote Sender's utility without communication, i.e., $V_0 = \mathbb{E}_{p_0} v(\hat{a}(p_0),\omega)$.

Generalizing Kamenica and Gentzkow's (2011) technique, an ambiguous communication device $(\pi_k)_K$ can also be expressed as an induced set of distributions over posteriors. Fixing p_0, each probabilistic device π_k induces $\mathbf{p}_k = (p_m^{\pi_k})_M$, the vector of posteriors at all messages, and $\tau_k \in \Delta M$, the marginal distribution over message space. Then the tuple $(\tau_k, \mathbf{p}_k)$ is a distribution over posteriors and by construction it is *Bayes plausible*, i.e., $\sum_{m\in M} \tau_k(m) p_m^{\pi_k} = p_0$. Let R denote the set of distributions over posteriors induced by these K probabilistic devices, i.e.,

$$\begin{aligned} R = \quad & \{(\tau_k, \mathbf{p}_k) \in \Delta M \times (\Delta\Omega)^M : \tau_k(\cdot) = \int_\Omega p_0(\omega')\pi_k(\cdot|\omega')d\omega', \mathbf{p}_k = (p_m^{\pi_k})_{m\in M}, \\ & supp(\tau_k) = supp(\tau_j), \forall \pi_k, \pi_j \in (\pi_k)_K\}. \end{aligned}$$

[15] 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.

[16] The exact label of the chosen probablistic device does not matter, since the ex ante Sender value is symmetric in labels. This tie-breaking rule is imposed to ensure the desired upper semi-continuity property of sender's interim value function.

[17] Note that with ambiguous beliefs and our assumption of updating maxmin EU preferences, Sender's ex ante and interim 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 decision-maker's preferences—the decision-maker 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 and, going backward, choose an ambiguous device according to Sender's ex ante MEU preferences.

Denote by $\mathcal{R}$ the collection of all such sets R, i.e., sets of distributions over posteriors induced by some ambiguous device.

We can rewrite Sender's problem as

$$\sup_{R\in\mathcal{R}} \min_{(\tau_k,\mathbf{p}_k)\in R} \mathbb{E}_{\tau_k}[\mathbb{E}_{p^k_m}[v(\hat{a}(P_m),\omega)|m]]. \tag{3}$$

Here $v_k(P_m) := \mathbb{E}_{p^k_m}[v(\hat{a}(P_m),\omega)|m]$ is Sender's utility when computed with respect to the k-th coordinate of the posterior vector $P_m = (p^1_m,\ldots,p^k_m)$, assuming Receiver best responds to posteriors P_m.[18]

4 A characterization of the value of persuasion

In this section, we present our main result—a characterization of the optimal value Sender can achieve by using K probabilistic devices. Geometrically, this value is given by the concave closure of the v_k function maximally projected to the first coordinate (See Figures 2–4). In proving this result, we introduce a splitting lemma that specifies the sets of plausible posteriors and a construction called synonyms that enables Sender to hedge against ex ante ambiguity. In addition, they provide a way to determine effortlessly the optimal ambiguous device in two steps. In the first step, we identify the optimal posteriors to elicit. The splitting lemma (Lemma 1) implies that the required posteriors are plausible. In the second step, we can construct synonyms by duplicating the messages leading to the optimal posteriors in such a way that Sender hedges against ex ante ambiguity.

4.1 A useful construction

We begin by characterizing the value function. In Kamenica and Gentzkow's (2011) model, 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 $v(p) = \mathbb{E}_p v(\hat{a}(p),\omega)$ the expected payoff of Sender when Receiver holds the Bayesian posterior belief $p \in \Delta(\Omega)$. In this subsection, we construct a similar "concavified" function for the case of ambiguous persuasion.

Let K be a finite number of Bayesian devices to be used. For any given vector of beliefs $P = (p_k)_K$, recall that $v_k(P) = \mathbb{E}_{p_k} v(\hat{a}(P),\omega)$ is the payoff of Sender when Receiver holds ambiguous posterior beliefs P and when computed with respect to the k-th coordinate in the posterior vector. By construction v_k is symmetric across devices,[19] without loss of generality we will focus on v_1.

[18] For simplicity of notation, we make no distinction between the vector of posteriors $P_m = (p^k_m)_K$ 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.

[19] Note that for all $k, j \in K$, $v_k(P) = v_j(P_{kj})$ where P_{kj} is the vector P with the k-th and j-th coordinates switched.

Recall that the subgraph of v_1 is $\{(P, z') \in (\Delta\Omega)^K \times \mathbb{R} | v_1(P) \geq z'\}$. Then, the concave closure of fuction v_1 is

$$V_1(P) = \sup\{z' | (P, z') \in co(\text{Subgraph}(v_1))\}.$$

By definition, if $z = V_1(P)$, then there exists a distribution of posteriors $\tau \in \Delta M$ and a (stacked) vector of posteriors $(P_m)_M \in \left((\Delta\Omega)^K\right)^M$ such that $\mathbb{E}_\tau P_m = P$ and $\mathbb{E}_\tau v_1(P_m) = z$.[20]

Consider now the maximal projection of $V_1(P)$ over $\Delta\Omega$, that is, let

$$\overline{V}(p) = \max_{P^{-1} \in (\Delta\Omega)^{K-1}} V_1(p, P^{-1}).$$

Our main result is Proposition 1. The first part says that the value of optimal ambiguous persuasion (with K devices) is determined by the above function. Sender benefits from ambiguous persuasion relative to Bayesian persuasion if and only if this value function is greater than the value function from Bayesian persuasion à la Kamenica and Gentzkow (2011). The second part says that only two probabilistic devices are needed to achieve this optimal value (while the number of messages M can vary as long as it is an finite integer).

Proposition 1. *Sender benefits from ambiguous persuasion if and only if $\overline{V}(p_0) > \hat{V}(p_0)$. Moreover, $\overline{V}(p_0)$ is independent from K as long as $K \geq 2$.*

Now, we adapt the illustrating example in Section 2 and demonstrate how to compute the $\overline{V}$ function and derive the value of ambiguous persuasion according to Proposition 1.

Example 2. We introduce a third action to the illustrating example in Section 2. Receiver now (the physician) has three possible choices: a competing generic drug (low action, a_l), a reputable old drug of the brand name drug producer (middle action, a_m), and a new and more profitable new drug by the same company (high action, a_h). As before, there are two possible states, high (ω_h) and low (ω_l), which reflects the effectiveness of the new drug relative to the competing generic drug. 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.

The payoffs are such that Receiver prefers the competing generic drug (a_l) if the probability of the high state is less than 1/4, and the new drug (a_h) if the probability is higher than 3/4, and safe old drug (a_m) for beliefs in between. The drug company's preference ranking is its new drug followed by its own old drug by the competitor's drug, which payoffs equal

[20] The maximum attains because V_1 is upper semi-continuous. To see this, apply Berge's theorem of maximum in each of the following steps: continuity of $u(a, P)$ implies the best response correspondence $a^*(P)$ is upper hemi-continuous; hence $v_1(P) = \max_{a \in a^*(P)} v(a, p_1)$ is also upper semi-continuous; this implies $V_1(P) = \sup_{\{\tau : \sum_m \tau(m) P_m = P\}} \tau(m) v_1(P_m)$ is upper semi-continuous.

Table 1: Payoff matrix for Example 2.

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

In each cell, the first number is Sender's payoff and the second is that of Receiver.

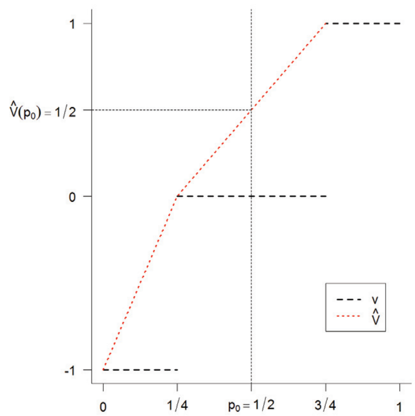

Figure 1: The value of Bayesian persuasion. The horizontal axis is the posterior probability of the low state and the vertical axis is Sender's payoff.

to $1, 0, -1$, respectively. The prior belief is $1/2$ and without further information the doctor will choose the safe old drug.

We first consider 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 the figure, he would want to generate two posteriors $1/4$ and $3/4$ with equal probabilities. In this case, he successfully persuades the doctor to prescribe the new brand name drug with probability $1/2$, and this is the optimal value of Bayesian persuasion, which yields 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 and prescribes the new brand name drug; when receiving the message "middle," Receiver is indifferent between a_m and a_l and takes the former (Sender-preferred) action and prescribes the old brand name drug.

We now turn to computing the value of ambiguous persuasion to Sender when two Bayesian devices are used, which, as we show in Proposition 1, is characterized by the function $\overline{V}$.

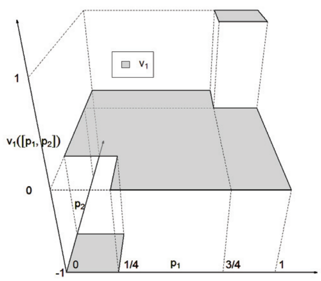

Figure 2: $v_1(P)$

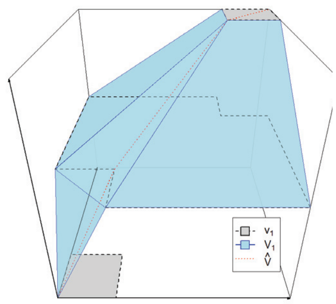

Figure 3: $V_1(P)$

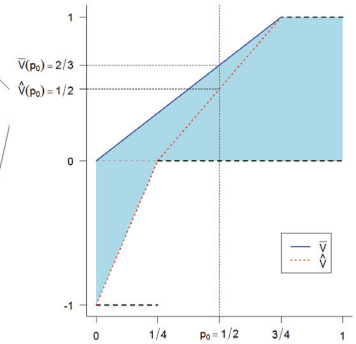

Figure 4: $\overline{V}(p)$

Figures 2, 3, and 4 illustrate the step-by-step construction of the $\overline{V}$ function. Firstly, with two (optimal) Bayesian devices used, the maxmin EU Receiver's (Sender-preferred) best response would then be a_h if both posteriors are at least $3/4$ and a_l if both posteriors are less than $1/4$, otherwise, she chooses the safe middle action a_m. Therefore, Sender's expected payoff based on his first posterior, as illustrated by Figure 2, is

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

Figure 3 then depicts V_1, the concave closure of v_1. Finally, Figure 4 illustrates function $\overline{V}$, the maximal projection of V_1 to the first dimension. The last figure depicts what Sender can achieve using two probabilistic devices. In this example, an optimal ambiguous device should lead to posterior sets $\{0, 1/4\}$ (when message "low" is received) and $\{3/4, 3/4\}$ (when message "high" is received). The first device would generate posteriors 0 and 3/4 and the second device will generate posteriors 1/4 and 3/4. Finally, this optimal ambiguous device yields value $\overline{V}(p_0) = 2/3$, which is higher than the optimal value of Bayesian persuasion. We give the precise construction of the ambiguous device in Section 4.3.

The full proof of Proposition 1 is given in appendix A.4. Our proof makes use of two intermediate results which we will introduce in the next two subsections.

4.2 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, the number of devices K, and the message space M, we look at the stacked vector/profile of posteriors $(p^k_m)_{M\times K} \in (\Delta\Omega)^{M\times K}$ and want to know whether it can be induced by some ambiguous communication device $(M, (\pi_k)_K)$ via full Bayesian updating.[21] Recall that a stacked vector of posteriors $(p^k_m)_{M\times K}$ is probability-by-probability updated by some ambiguous communication device, if there exists some device $(M, (\pi_k)_K)$ such that for all $m \in M$ and $k \in K$,

$$p^{\pi_k}_m(\cdot) = \frac{\pi_k(m|\cdot)p_0(\cdot)}{\int_\Omega \pi_k(m|\omega')p_0(\omega')d\omega'}$$

is just the posterior of Bayesian device π_k. As usual, we require the ambiguous device $\Pi = (\pi_k)_K$ to have common support on M; i.e., $\tau^{\pi_k}(m) > 0 \Rightarrow \tau^{\pi_j}(m) > 0$ for all $k, j \in K$, where $\tau^{\pi_k}(m) = \int p^{\pi_k}_m(\omega')p_0(\omega')d\omega'$.

Definition 1. We say a stacked vector of posteriors $(p^k_m)_{M\times K} \in (\Delta\Omega)^{M\times K}$ is *potentially generalized Bayes plausible* (PGBP) if for all $k \in K$ there exists some $\tau^k \in \Delta M$ such that $p_0 = \sum_m \tau^k(m)p^k_m$, with $supp(\tau^k) = supp(\tau^j)$ for all $k, j \in K$.

The notion of PGBP requires that posteriors have to be Bayes plausible device by device (although the weights on the posteriors could differ across devices).

Our next lemma shows full Bayesian updating via some *unobserved* ambiguous device is equivalent to the potentially generalized Bayes plausibility condition. 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, K, and message space M. A stacked vector of posteriors $(p^k_m)_{M\times K}$ is potentially generalized Bayes plausible if and only if $(p^k_m)_{M\times K}$ consists of the probability-by-probability updated posteriors via some ambiguous device $(M, (\pi_k)_K)$ with common support.*

Lemma 1 implies that not every vector of posteriors could correspond to the updated posteriors by some ambiguous communication device. It puts a bound on the vector 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 plausibility under one probabilistic device, as the posterior weights τ^k could differ across devices, but this freedom is not without limit. This is illustrated by the following example.

Example 3. Consider for instance Example 2. Just like Kamenica and Gentzkow's (2011) construction, ours is useful to determine sets of posteriors that we would want to generate from an ambiguous device. In this case, two messages are needed, which we denote by m_l and m_h. It follows from Figure 4 that the first device would need to yield posteriors of

[21] By considering the vector of posteriors conditional on a message m, we keep track of which probabilistic device each posterior is updated from.

$p_{1,m_l} = 0$ and $p_{1,m_h} = 3/4$. By construction, these are necessarily Bayes plausible. 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, posteriors in $\arg\max_p v_k(\{p_{1,m_l}, p\})$ and $\arg\max_p v_k(\{p_{1,m_h}, p\})$.

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	1/4	3/4

$p_2(\omega\|m)$	ω_l	ω_h
m_l	3/4	1/4
m_h	1/4	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	2/3	0
m_h	1/3	1

$\pi_2(m\|\omega)$	ω_l	ω_h
m_l	3/4	1/4
m_h	1/4	3/4

Furthermore, note that if Sender were to use the ambiguous device Π derived from the two Bayesian devices above, then receiver would choose a_h upon seeing m_h and a_m upon seeing m_l. Note also that Bayesian device 1 sends the two messages with probabilities $\tau_1 = (2/3, 1/3)$. So, we have that $v_1(p_1, p_2) = 2/3 = \overline{V}(p_0)$.

Note that, however, 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 the profile of posteriors of the first probabilistic device is, as remarked above, necessarily Bayes-plausible. Only those posterior profile resulting from the maximal 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	1/4	3/4
m_r	7/8	1/8

$p_2(\omega\|m)$	ω_l	ω_h
m_l	3/4	1/4
m_h	1/4	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	$20(1-\varepsilon)/21$	0
m_h	$(1-\varepsilon)/21$	$1-\varepsilon$
m_r	ε	ε

$\pi_2(m\|\omega)$	ω_l	ω_h
m_l	3/28	1/4
m_h	1/28	3/4
m_r	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 $\lim_{\varepsilon\to 0} \tau_1 = (5/6, 1/6, 0)$ and $\lim_{\varepsilon\to 0} v_1(p_1, p_2) = 5/6*(0)+1/6*1 = 1/6 = \overline{V}(1/8)$.

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

4.3 The use of synonyms

In this section, we consider messages that are *(strong) synonyms*, which are multiple messages that yield the same set of posterior beliefs. We first introduce a way to construct strong synonyms from any ambiguous communication device by duplicating the message space. We then show that Sender can hedge himself against ambiguity by constructing messages in such a manner.

First, we 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.

Consider now an ambiguous device $\Pi = co((\pi_1, \pi_2))$. 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 strong 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.

Suppose now Sender considers such a probabilistic device π_1' while Receiver still best responds to posterior beliefs by ambiguous device Π'. Let $V(\pi_1')$ be the value of such a device. We define $V(\pi_1)$ likewise. 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. See Subsection A.2. □

Using Lemma 2, the value of Π' is $V(\Pi') = \alpha V(\pi_1) + (1-\alpha)V(\pi_2)$. Without loss assuming 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. Also $V(\pi_1')$ and $V(\pi_2')$ both approximate the same value of $\max\{V(\pi_1), V(\pi_2)\}$ as $\alpha \to 1$. From this we have the following lemma.

Lemma 3. *Given an ambiguous device $\Pi = co((\pi_k)_K)$, there exists a sequence of devices $(\Pi_n')_{n\in\mathbb{N}} = \{co(\pi_{k,n}')_K\}_{n\in\mathbb{N}}$ using synonyms such that $\underset{n\to+\infty}{lim} V(\Pi_n') = \underset{\pi\in\Pi}{sup} V(\pi)$. Moreover, $\lim_{n\to\infty} V(\pi_{k,n}') = \underset{\pi\in\Pi}{sup} V(\pi)$ for all k.*

Proof. See Subsection A.3 for a proof for the general case (more than two probabilistic devices). □

Observe that in the sequence of ambiguous devices with synonym construction, the values of all probabilistic devices converge to the same limit. At an optimal ambiguous device, Sender must be indifferent among all the probabilistic devices used; otherwise he could always use synonyms to attain the highest value.

Below, we continue with 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	2/3	0
m_h	1/3	1

$\pi_2(m\|\omega)$	ω_l	ω_h
m_l	3/4	1/4
m_h	1/4	3/4

These devices lead to the following distributions over posteriors:

	$\tau_1(m)$	$p_1(\omega_h\|m)$
m_l	1/3	0
m_h	2/3	3/4

	$\tau_2(m)$	$p_2(\omega_h\|m)$
m_l	1/2	1/4
m_h	1/2	3/4

Recall the value of the mentioned optimal ambiguous device is $\overline{V}(1/2) = 2/3$. 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 $V(\pi_1) = 2/3$; while the value when computed with regard to π_2, $V(\pi_2)$, is however 1/2. Sender's maxmin value of the ambiguous device Π is therefore only 1/2.

Nevertheless, by Lemma 3, one can then use synonyms 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	$\alpha \cdot 2/3$	0
m_h	$\alpha \cdot 1/3$	α
m'_l	$(1-\alpha) \cdot 3/4$	$(1-\alpha) \cdot 1/4$
m'_h	$(1-\alpha) \cdot 1/4$	$(1-\alpha) \cdot 3/4$

$\pi'_2(m\|\omega)$	ω_l	ω_h
m_l	$(1-\alpha) \cdot 3/4$	$(1-\alpha) \cdot 1/4$
m_h	$(1-\alpha) \cdot 1/4$	$(1-\alpha) \cdot 3/4$
m'_l	$\alpha \cdot 2/3$	0
m'_h	$\alpha \cdot 1/3$	α

These devices lead to the following distributions over posteriors:

	$\tau'_1(m)$	$p_1(\omega_h\|m)$
m_l	$\alpha \cdot 1/3$	0
m_h	$\alpha \cdot 2/3$	3/4
m'_l	$(1-\alpha) \cdot 1/2$	1/4
m'_h	$(1-\alpha) \cdot 1/2$	3/4

	$\tau'_2(m)$	$p_2(\omega_h\|m)$
m_l	$(1-\alpha) \cdot 1/2$	1/4
m_h	$(1-\alpha) \cdot 1/2$	3/4
m'_l	$\alpha \cdot 1/3$	0
m'_h	$\alpha \cdot 2/3$	3/4

In this case, Receiver will best respond with a_m upon seeing m_l or m_l', which takes place with (joint) probability $1/2 - \alpha/6$ under either π_1' and π_2'; and best respond with a_h upon seeing m_h or m_h', occuring with probability $\alpha 2/3 + (1-\alpha)1/2$ under either 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 Lemma 3 in the following manner:

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

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

The induced distributions over posteriors are:

	$\tau_1'(m)$	$p_1(\omega_h\|m)$
m_l	$\alpha \cdot 5\varepsilon/6$	0
m_h	$\alpha \cdot \varepsilon/6$	$3/4$
m_r	$\alpha(1-\varepsilon)$	$1/8$
m_l'	$(1-\alpha)\cdot 1/8$	$1/4$
m_h'	$(1-\alpha)\cdot 1/8$	$3/4$
m_r'	$(1-\alpha)\cdot 3/4$	0

	$\tau_2'(m)$	$p_2(\omega_h\|m)$
m_l	$(1-\alpha)\cdot 1/8$	$1/4$
m_h	$(1-\alpha)\cdot 1/8$	$3/4$
m_r	$(1-\alpha)\cdot 3/4$	0
m_l'	$\alpha \cdot 5\varepsilon/6$	0
m_h'	$\alpha \cdot \varepsilon/6$	$3/4$
m_r'	$\alpha(1-\varepsilon)$	$1/8$

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 $(\alpha\varepsilon \cdot 1/6 - 5/8 \cdot (1-\alpha) - \alpha(1-\varepsilon))$ which converges to $\overline{V}(1/8) = 1/6$ when α and ε converge to 1.

Finally we observe that the value of ambiguous persuasion is the same regardless of the number of Bayesian devices used (so long as two are used).

Corollary 1. *For every ambiguous device $(\pi_k)_K$ there is an ambiguous device with only two probabilistic devices and attains the same value.*

Proof. Let $\Pi = (\pi_k)_K$ be an ambiguous device of value V. By Lemma 3, it is without loss to assume that the value of Π when computed from any of the K probabislitic devices, $V(\pi_k)$, is the same. For all m, there exists[22] $\tilde{p}_m \in P_m$ such that $\hat{a}(\{\tilde{p}_m\}) = \hat{a}(P_m) = \hat{a}([\tilde{p}_m, p_1])$. The rest of the construction mimics that of the characterization proof. □

[22]This is a simple application of Kakutani's fixed point theorem and is a known result.

4.4 The necessity of synonyms

In the previous subsection, we have shown how Sender may use synonyms to approach the highest value among several probabilistic devices. In this subsection, we investigate whether synonyms are necessary for optimal ambiguous persuasion to be beneficial and prove that this is generically true.

We start by introducing a distinction between *strong synonyms* and *weak synonyms*. So far, we have focused on *strong synonyms*, which, as previously mentioned, are multiple messages that lead Receiver to hold the same *set of posterior beliefs*. In contrast, multiple messages are said to be *weak synonyms* if they lead Receiver to take the same *action*. Note that all strong synonyms are also weak synonyms but the converse is not true. Furthermore, strong synonyms are a characteristic of the communication device and are immune to changes in Receiver's payoffs whereas weak synonyms are a characteristic of both the communication device and Receiver's payoffs.

In this subsection, we show that under two technical assumptions imposed to rule out corner cases, if ambiguous persuasion is beneficial (compared to Bayesian persuasion), then the optimal ambiguous device uses weak synonyms.[23] That is, an ambiguous device that is both optimal and beneficial cannot be *straightforward* à la Kamenica and Gentzkow (2011).[24] Since for every Bayesian device there exists an equivalent straightforward device, the (generic) necessity of weak synonyms distinguishes ambiguous persuasion from Bayesian persuasion.

The result may not however hold in some special cases. For instance, if two actions are so similar to the point of redundancy, then messages leading to these actions could *de facto* perform the same role as two weak synonyms. However, in these cases, small perturbations of payoffs would break the pattern.

Before ruling out these corner cases with two assumptions, we introduce the following notation.

For an arbitrary k, let $\overline{v}$ be the maximal projection of v_k. By construction, note that $\overline{V}$ can also be defined as the concave closure of $\overline{v}$. Indeed, $\overline{V}$ is defined as the maximal projection of the convex hull of v_k. Given both steps are a maximization, it does not matter in which order these two steps are taken.

Let H_0 be a supporting hyperplane of the subgraph of function $\overline{V}$ at point $(p_0, \overline{V}(p_0))$,

[23] In order to state there exists an "optimal" ambiguous device, we may assume that when a probabilistic device never sends a message in expectation, the posterior resulting from this device at the given message can be chosen arbitrarily. This can be argued as one could construct a sequence of devices that have this posterior at said message with the probability that this message is sent converging to 0.

[24] A device is said to be straightforward if each message leads to a single different action, that is, if it does not use weak synonyms.

defined by

$$H_0 = \{(p, v) \in \Delta\Omega \times \mathbb{R} : \langle p, \xi \rangle - v = \langle p_0, \xi \rangle - \overline{V}(p_0), \text{ where } \xi \in \partial \overline{V}(p_0)\}.$$

Such a hyperplane exists as $\overline{V}$ is concave by construction (and its subgraph is a convex set) and $(p_0, \overline{V}(p_0))$ is included in *subgraph*$(\overline{V})$. Note that such a hyperplane is unique if and only if $\overline{V}$ is differentiable at p_0. Our first assumption relates to the uniqueness of this hyperplane.

Assumption 1. $\overline{V}$ is differentiable at p_0.

To see why Assumption 1 is not restrictive for us, first observe that it is satisfied in all the examples discussed in this paper. Moreover, note that if $\overline{v}(p_0) < \overline{V}(p_0)$, then it must be that $\overline{V}$ is linear at p_0. Thus differentiability is satisified whenever the optimal ambiguous device is valuable.

We now introduce a new concept of a *relevant* posterior. We say that a posterior is *relevant* if $(p, \overline{v}(p)) \in H_0$. Lemma 4 below will show that, for an optimal device that yields value $\overline{V}(p_0)$, any posterior that may be induced with non-zero probability must be revelant. Furthermore, by the definition of $\overline{v}$, for any posterior p, there exists an action a such that $\mathbb{E}_p v(a, \omega) = \overline{v}(p)$ and $a = \hat{a}(P)$ for some posterior set P that includes p. We denote $\overline{a}(p)$ the set of such actions. We say that an *action-posterior pair* (a, p) is *relevant* if p is relevant and $a \in \overline{a}(p)$. Relevant action-posterior pairs are of interest to us as those are the only ones that can arise in equilibrium with non-zero probability for any of the Bayesian devices.

Lemma 4. *Let $\Pi = (\pi_k)_K$ be an optimal ambiguous device. If a message is sent with non-zero probability from one of the Bayesian devices π_k, then the action taken by Receiver and the posterior belief associated with the Bayesian device at said message form a relevant action-posterior pair.*

Proof. See Subsection A.5. □

The intuition of the result is that if the action and posterior does not form a relevant action-posterior pair, then the resulting expected payoff of Sender will be strictly lower than $\overline{V}(p_0)$, contradicting the fact that Π is an optimal ambiguous device.

We can now state our second assumption that rules out corner cases.

Assumption 2. If an action-posterior pair $(\hat{a}, \hat{p})$ is relevant, then for sufficiently small $\varepsilon > 0$ and all perturbation $(\eta)_\omega \in [-\varepsilon, \varepsilon]^\Omega$ that modifies payoff of $\hat{a}$ to $v^\varepsilon(\hat{a}, \omega) = v(\hat{a}, \omega) + \eta_\omega$ such that H_0 remains unchanged, $(\hat{a}, \hat{p})$ must also be relevant for the perturbed game.

Assumption 2 requires that relevant action-posterior pairs must still be relevant for small perturbations to the actions' payoffs to Sender. Figure 5 illustrates how this assumption is satisfied in our running example, and Figure 6 illustrates in three cases how it may fail.

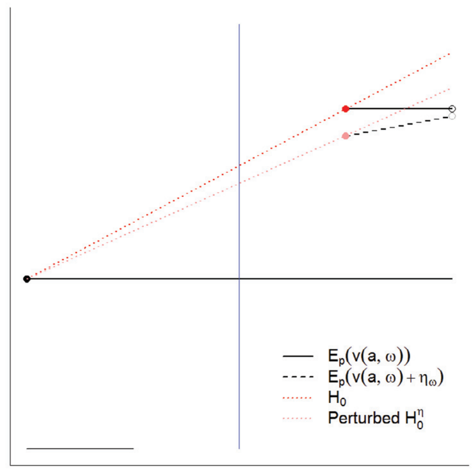

Figure 5: An action-posterior pair that remains relevant after a perturbation. The full dots correspond to relevant action-posterior pairs for the original $\overline{v}$ function in solid black line. The black dashed line corresponds to a small perturbation of Sender's payoffs which moves the hyperplane H_0 downward from the bright red dotted line to the dim red dotted line. The bright red dot—the action-posterior pair that is relevant in the original game—is lowered to the dim red dot in the perturbed game. The latter is still relevant as it is on the perturbed hyperplane H_0.

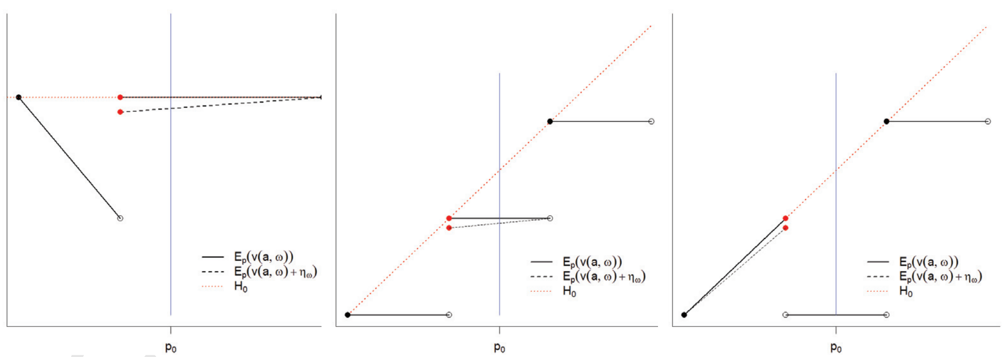

Figure 6: Three examples where Assumption 2 is not satisfied. In each case, the red dot corresponds to an action-posterior pair that is relevant in the original game but no longer so after a small perturbation of payoffs illustrated by the dashed line. In each case the perturbation was chosen so that $\overline{V}$ and therefore H_0 was left constant.

Proposition 2. *Suppose Assumptions 1 and 2 hold. Then, either there is no value to ambiguous persuasion (beyond that of Bayesian persuasion), or the optimal ambiguous device uses weak synonyms.*

Proof. See Subsection A.6. □

A sketch of the proof for the case of two devices is as follows. Suppose $\Pi = \{\pi_1, \pi_2\}$ is an optimal straightforward ambiguous device that does not use weak synonyms. Then, Assumption 2 implies that the set of action-posterior pairs occurring with non-zero probability must be the same for both devices. To see this, assume it is not the case and there can be two types of violations: first, there is an action a that occurs with positive probability under device π_1 (paired with relevant posterior p) but occurs with probability zero under device π_2; second, an action played with non-zero probability is associated with different relevant posteriors induced by the two devices. To rule out the first type of violation, one could introduce a small perturation to Sender's value at action a in the manner of $\overline{v}^{\varepsilon}(a, \cdot) = \overline{v}(a, \cdot) - \varepsilon$ without affecting sender's payoff at other actions. Since action a is played with probability zero under the second device, in the perturbed game the value of the second device and hence the optimal value from the two devices remain the same. By straightforwardness, the values of $\overline{v}$ at all other relevant posteriors associated with other messages that are sent with non-zero probabilities remain the same, and so is the value $h_0(p)$, where $h_0(\cdot)$ is defined such that $(p, h_0(p)) \in H_0$. Yet $\bar{v}^{\varepsilon}(p) < \bar{v}(p)$, implying that $(p, \bar{v}^{\varepsilon}(p))$ falls below the hyperplane H_0 and hence (a, p) is no longer a relevant pair, which is a contradiction. Therefore, the set of actions occurring with non-zero probability in equilibrium must be the same for the two devices. The second type of violation can also be ruled out by a similar perturbation argument. Combining the previous observation and the fact that in a straightforward device only one message can be associated with a relevant action-posterior pair, we have the posterior set at each message must be a singleton. Therefore, Sender can build a Bayesian device that leads to the same actions by Receiver and yields the same value—ambiguous persuasion has no extra value.

The main reason why (weak) synonyms can solve the conundrum exposed in the proof is the fact that allowing weak synonyms lets π_1 and π_2 use two different messages to lead to the same relevant action-posterior pair.

5 Examples

5.1 When Sender-optimal Action Is Receiver-safe

In the introductory example (Section 2), Sender has an optimal action—prescribing the brand name drug—that is also a safe default for Receiver. In this case, Sender can benefit from sending two ambiguous messages that are synonyms, as both messages generate the same set of posteriors. In this section, we generalize the intuition to the case with finitely many states and actions.

We first introduce a few assumptions.

Assumption 3. Sender has a state-independent optimal action, i.e., there is some $\hat{a} \in A$ such that $v(\hat{a}, \omega) \geq v(a, \omega)$ for all $a \in A$ and $\omega \in \Omega$.

Assumption 4. $\min_{\omega} u(\hat{a}, \omega) \geq \min_{\omega} u(a, \omega)$ for all $a \in A$.

Assumption 5. Receiver will not choose $\hat{a}$ without persuasion, i.e., there is some $a \in A$ such that $\mathbb{E}_{p_0}[u(a, \omega)] > \mathbb{E}_{p_0}[u(\hat{a}, \omega)]$.

A familiar special case of Assumption 3 is that Sender's preferences are state-independent, $v(a, \omega) = v(a)$ for all ω. Assumption 4 requires action $\hat{a}$ is also the "safest" for Receiver in the sense that it yields the highest worst-case payoff. To focus on the interesting case, Assumption 5 just requires with only prior information, Receiver will not choose $\hat{a}$.

Proposition 3. *Under Assumptions 3–4, there exists an optimal ambiguous communication device Π^* that ensures that Receiver always take the Sender optimal action $\hat{a}$.*

Under Assumptions 3–5, Sender will do strictly better with Π^* than with any Bayesian communication device.

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 equal 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 4, for all $a \neq \hat{a}$, there is some ω_a such that

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

and hence Receiver will choose $\hat{a}$ at all message $m \in M$. Under Assumption 3, this communication device (M, Π) is Sender optimal.

Now, we show that, under Assumption 5, 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 $\hat{a}$ is not an optimal action for Receiver and hence not chosen by her. Suppose instead action $\hat{a}$ is chosen under every posterior p_m such that $\tau(m) > 0$, that is,

$$\mathbb{E}_{p_m} u(\hat{a}, \omega) \geq \mathbb{E}_{p_m} u(a, \omega) \qquad \text{for all } a.$$

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

$$\mathbb{E}_{p_0} u(\hat{a}, \omega) = \sum_m \tau(m) \mathbb{E}_{p_m} u(\hat{a}, \omega) \geq \sum_m \tau(m) \mathbb{E}_{p_m} u(a, \omega) = \mathbb{E}_{p_0} u(a, \omega)$$

for all a, contradicting Assumption 5. □

In this example, Sender hedges against any ex ante ambiguity by sending $|M| = |\Omega|$ messages that are strong synonyms to begin with—they generate the same posterior set featuring maximal ambiguity, i.e., $P_m = \Delta(\Omega)$. In response, an ambiguity-averse Receiver will take the Sender preferred action $\hat{a}$ at all messages, because it is also the safest choice for her. Sender succeeds in persuasion by muddying the waters.

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.

As Crawford and Sobel (1982) remark, it is straightforward to show that the optimal probabilisitc communication device for Sender is to fully disclose ω to Receiver.[25] To see this, let m be the realized message, then Receiver would take an action

$$a^*(m) = \mathbb{E}(\omega|m) + b$$

upon observing m. Also, $\hat{a}(m) = a^*(m)$ as the best response is unique. Thus, Sender must choose a communication device to solve the maximization problem

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

which becomes

$$E\left[\mathbb{E}(\omega|m)-\omega\right]^2 + b^2,$$

and minimized when

$$\mathbb{E}(\omega|m) = \omega.$$

Now, we investigate whether it is possible for Sender to do strictly better using ambiguous devices, even if Sender is ambiguity averse himself.

[25]Throughout the paper we assume that M is finite. To reconcile this assumption with the optimal probabilistic device here, consider a finite but large message space $M = \{m_1, m_2, \ldots, m_N\}$, where each m_n maps to a posterior uniformly distributed on the interval $[(n-1)/N, n/N]$. The Sender's value under optimal Bayesian persuasion can be approximated by letting $N \to +\infty$. In the case of the optimal simple ambiguous device, M is finite.

We focus our attention on *simple ambiguous communication devices*, which are characterized by a message set $M = \{m_{1A}, m_{1B}, \ldots, m_{nA}, m_{nB}\}$, a partition of the $[0, 1]$ interval with $2n$ 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 $y_n = 1$,[26] as well as probabilistic devices $\Pi(\mathbf{y}, \mathbf{c}) = \{\pi, \pi'\}$. 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 $i = 1, \ldots, n$. The probabilistic devices π and π' we consider are of the form

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 $Uni(I_{iA})$ and $Uni(I_{iB})$ refer to the uniform distributions on I_{iA} and I_{iB}, respectively. In other words, messages m_{iA} and m_{iB} are strong synonyms. Our construction borrows from Kellner and Le Quement (2018), who show that using the above devices there exist Pareto-improving *cheap-talk* equilibria when Sender is allowed to send ambiguous messages.

To characterize the optimal simple ambiguous communication device, we first compute Receiver's optimal action 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 we may focus on the simplified case below and use translation to obtain results for $\{m_{iA}, m_{iB}\}$ and $\{I_{iA}, I_{iB}\}$. Let $I = [0, l)$, $I_A = [0, c]$, and $I_B = (c, l)$ for $c \in I$ and let the signals π and π' be analogously defined as above. The following lemma characterizes Sender's optimal choice of cutoff c.

Lemma 5. *Sender's optimal cutoff is*

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

and correspondingly, Receiver's optimal action is

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

[26] In this definition, to be rigorous, we have another partition cell $[y_n, y_n] = \{1\}$. Alternatively, we could also make $I_{nB} = (y_{n-1} + c, y_n] = (y_{n-1} + c, 1]$. However, since it does not impact payoffs of either player, we leave out any discussion about what occurs at $y_n = 1$.

In the uniform-quadratic case, Receiver always has a unique optimal response $a^*(l,b)$, which is the same as $\hat{a}(l,b)$.

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 6. *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 4. *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.*

The optimal information structure we characterize in Proposition 4 satisfies both *equal intervals* and *maximum ambiguity*, in the terminology of Kellner and Le Quement (2018) (KLQ hereafter).[27] 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 an endpoint of the interval. Given our assumption that Sender has full commitment, unlike KLQ, we place no incentive compatibility requirement on Sender.[28] In our characterized optimal information structure, the state space $[0,1]$ is divided

[27] 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.

[28] Since KLQ focus on cheap-talk communication, Sender's incentive compatibility constraints have to be satisfied. Thus, equal intervals and maximum ambiguity cannot both hold unless $l = 6b$, which is only possible if $6b = 1/n$ for some integer n. Furthermore, it is not necessarily that this equilibrium is the most informative when this is the case. See Appendix A.10 for a detailed analysis.

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, the best Sender could do is to completely reveal all the information to Receiver. However, with such possibility, Sender faces a tradeoff between being precise 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 Sender's ideal action. Consequently, committing to being fully precise is no longer desirable, as is the case under expected utility, and instead Sender finds it optimal to maintain some degree of flexibility of communication devices and hence their interpretations.

6 Discussion

6.1 Dynamic Consistency

It is well known that with full Bayesian updating, ambiguity averse decision makers can be dynamically inconsistent (Epstein and Schneider, 2003). In this section, we consider the set of ambiguous communication devices that induce consistent behavior from any receiver. We 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. By choosing a dynamically consistent ambiguous device Sender can make sure that it is accepted by any receiver, which is useful if Sender is speaking in public to a group of receivers with different payoffs or if he is uncertain about the payoff of the particular receiver he is communicating to. Yet we show that such a set of dynamically consistent devices, though not empty, does not allow the expert to benefit from ambiguous persuasion.

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

Definition 2. Communication device Π is *dynamically consistent* if for all $\mathbf{a} \in A^M$ and $u : A \times \Omega \to \mathbb{R}$ and for all m, $U(a_{1m}, P_m) \geq U(a_{2m}, P_m) \Rightarrow U(\mathbf{a}_1, \Pi) \geq U(\mathbf{a}_2, \Pi)$.

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 $\Omega \times M$.[29] We adapt here their definition to our frame-

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

work. 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 3. 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 7. 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 $p = \pi(m) \cdot p_{\pi'}(\cdot|m) + \sum\limits_{m' \neq m} \pi(m') \cdot p_\pi(\cdot|m')$ 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 2. Π *is dynamically consistent if and only if* Π *is rectangular.*

Proof. See Subsection A.11. □

If the expert is restricted to dynamically consistent devices, then, from Lemma 7, 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 to another. Nevertheless, any action that such a device would entail could have been induced by a risky device as well. Unsurprisingly then, Sender cannot benefit from ambiguous persuasion if he is restricted to rectangular devices.

Proposition 5. *If Sender is restricted to dynamically consistent/rectangular devices, then there are no gains to playing ambiguous strategies.*

Proof. Let $\pi \in \arg\min\limits_{\pi \in \Pi^*} V(\pi, \hat{\mathbf{a}})$. Then playing the probabilistic strategy π induces the same strategy of Receiver (given rectangularity implies singleton posteriors) and thus $V(\{\pi\}) = V(\Pi^*)$. □

Note that the restriction to dynamically consistent devices is a strong one, which entails two conditions. First, it imposes a constraint on Receiver's preferences over actions that she does not take in equilibrium. Second, it imposes dynamic consistency not only on the strategy of Receiver who faces Sender, but also on that of *any* receiver.

Our next subsection looks at devices that result from mild relaxation of both conditions, while still ensuring the particular Receiver has consistent preferences in equilibrium. We show that they are not valuable to Sender. Then, we show more substantial weakening of either of the two conditions would allow the expert to benefit from ambiguous persuasion. Subsection 6.3 relaxes the first condition, while Subsection 6.4 relaxes the second.

6.2 Weak dynamic consistency

In this section, we look at a weaker condition, which only requires the particular Receiver's preferences over actions to be consistent on the equilibrium 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 $\hat{a}_m = \hat{a}(P_m)$ 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 $\hat{a}_m$, i.e.,

$$Q^*_m := \{q_m \in co(P_m) : \mathbb{E}_{q_m}[u(\hat{a}_m, \omega)] \geq \mathbb{E}_{q_m}[u(a, \omega)] \qquad \forall a \in A\}.$$

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 Section 2's 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 $\hat{a}_m$ =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 4. The ambiguous device Π is *semi-rectangular* if its induced set of distributions over posterior satisfies $R = \bar{R}$.

The next proposition says that for a semi-rectangular ambiguous device to be valuable, there must not be an "equivalent" Bayes plausible distribution of posterior supporting the same receiver strategies. Otherwise, a Bayesian device could mimic the same receiver responses and be more valuable for Sender.

Proposition 6. *If an ambiguous communication device* Π *is semi-rectangular and there exists a potentially Bayes plausible selection from* $(Q^*_m)_M$*, then the device* Π *has no value beyond Bayesian persuasion.*

Proof. See Subsection A.12. □

Note that both conditions are weaker than dynamic consistency. Semi-rectangularity requires the set of distributions of posteriors R to include all plausible mixtures of its projected posteriors. The other condition on the existence of a potentially Bayes plausible profile of posteriors is implied by rectangularity and it suggests that the same profile of receiver actions can be induced by the ambiguous device and some probabilistic device. This condition is clearly violated in our introductory example, as $Q^*_m = \{(p, 1-p) : p \in [0, 1 - \frac{c}{u_H - u_L}]\}$ for $m = e, i$ but $1 - \frac{c}{u_H - u_L} < p_0$, and hence $(Q^*_m)_M$ is not potentially generalized Bayes plausible.

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

Schlee (1997) shows 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 persuasion that would be valuable to any receiver.

Consider for example the following payoffs, where *a priori* the two states ω_l and ω_h are equally likely.

	ω_l	ω_h
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_1, 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 Sender 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 $\Pi = co((\pi_1, \pi_2))$. 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 Sender could benefit from ambiguous persuasion 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 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.13 proves more rigorously a (slightly) more general result[30].

Note that although Π' is valuable, it is not proven yet that Sender 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 Receiver and Sender are both linear in the $\oplus$ operation, it follows that valuableness is preserved when using synonyms as in Lemma 3. Thus, Sender can use synonyms to capture, with a valuable ambiguous device, the value of $5/9 > 1/2$.

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

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 Sender to modify an ambiguous device in order to satisfy a particular receiver's participation constraint: $U(\Pi) \geq U_0$ for a given u. In this case, Receiver would rather listen to the chosen ambiguous device Π than no listen at all.

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 7. *If Sender benefits from ambiguous persuasion (if $\bar{V}(p_0) > \hat{V}(p_0)$) and Receiver benefits from the optimal Bayesian persuasion, then there exists some ambiguous device that benefits Sender and satisfies Receiver's participation constraint.*

Proof. Let Π be the optimal ambiguous device and π be the optimal probabilistic 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, 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 $U(\Pi') \geq U_0$. $\square$

Consider for example the game represented by the following Figure 7 that provides the $\hat{v}$ function. In this case, Receiver would benefit strictly from probabilisitc persuasion compared to no communication. Thus, if it happens that Sender benefits from ambiguous persuasion[31], it is possible to find an ambiguous device that benefits Sender while Receiver is willing to listen.

A second method, which is illustrated by the second example, is to restrict Sender to "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 Receiver) if $u(\hat{a}(P_m), p_m) \geq u(a_0, p_m)$ for all posteriors in P_m. The following characterization applies:

[31] 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 persuasion would have no benefit here is if the safest action was the low action.

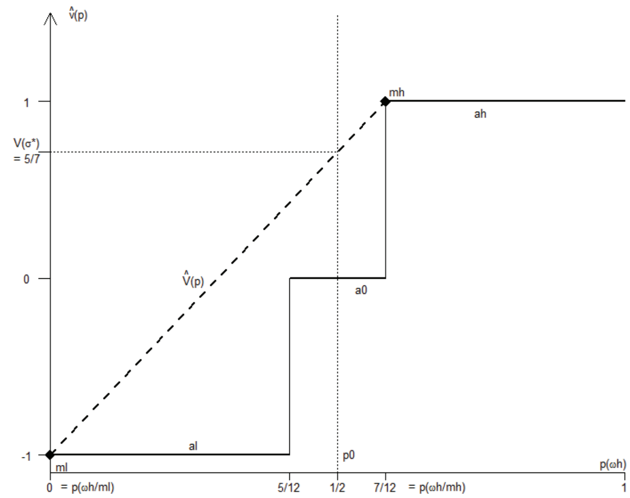

Figure 7: The value of ambiguous persuasion

Proposition 8. *A message is value-increasing (to 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 Receiver benefits from listening to device* Π*.*

Proof. See Section A.14. □

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 directly would ensure that the resulting ambiguous device satisfies the participation constraint of Receiver.

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

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

6.5 Other ambiguity preferences

In this paper, the set of beliefs is interpreted as "objective ambiguity" created by the device that Sender designs. Hence, assuming maxmin EU equates the objective and subjective sets of beliefs, and thus may be considered too extreme by some. Therefore, it is natural to ask whether our main characterization is robust to a wider range of ambiguity sensitivity with regard to Sender's and especially Receiver's preferences. We explore these questions in the subsection.

We first consider a wide class of ambiguity preferences for Sender. For a given ambiguous device $\Pi = co((\pi_k)_K)$, let $\mathbf{v} = \{\mathbb{E}_{p_0}[\mathbb{E}_{\pi_k}[v(\hat{a}(P_m), \omega)|\omega]]\}_K$ be the vector of Sender-expected utility for each individual probabilistic devices. We only impose the assumption that Sender's preferences are such that the value of a device Π must be between the minimum and the maximum value of probabilistic devices. That is, for any ambiguous device Π, $\min_{v \in \mathbf{v}} v \leq V(\Pi) \leq \max_{v \in \mathbf{v}} v$ where $\mathbf{v}$ is the vector of expected utilities as defined previously. We call it the *Betweenness* assumption.[32]

Corollary 3. *If Sender's preferences satisfy Betweenness, then Sender benefits from ambiguous persuasion if and only if* $\overline{V}(p_0) > \hat{V}(p_0)$.

Proof. See Appendix A.15. □

Intuitively, this result stems directly from the fact that Sender can perfectly hedge against ambiguity with synonyms. As a result, his exact preferences regarding ambiguity are irrelevant to the value of ambiguous persuasion.

Consider now Receiver's preferences. Receiver's preferences only matter insofar as they determine which action Receiver would take at any given posterior set. As a result, so long as the construction of function $\overline{V}$ reflects the actions actually taken by Receiver, the characterization result does not depend on Receiver's form of ambiguity preferences.

The only assumption needed here is that Receiver's choice of action depends only on the posterior set of beliefs as obtained from full Bayesian updating.

Corollary 4. *If Receiver's preferences at any posterior set* $U(a, P_m)$ *only depend on the full set* P_m*, then Sender benefits from ambiguous persuasions if and only if* $\overline{V}_U(p_0) > \hat{V}(p_0)$.

[32] The betweenness assumption holds in many utility representations in the literature. For example, it covers the α-min representation as in Gajdos et al. (2008), the α-MaxMin as in Jaffray and Philippe (1997), Choquet-Expected utility with convex or concave capacity of Schmeidler (1989) (with the core of the capacity or the core of its dual included in the belief set P), subjective expected utility of Anscombe and Aumann (1963) (with the subjective belief included in P), variational preferences of Maccheroni et al. (2006) (with the domain of the cost function included in P), and the smooth ambiguity preferences of Klibanoff et al. (2005) (with the support of the the second-order belief included in P).

Proof. Note that Receiver's preferences only influence $\overline{V}_U(p_0)$ via her best response function $\hat{a}_U(P_m)$. We can modify Sender's interim expected utility according to device k to $v_k(P_m) = E_{p_{m,k}}[v(\hat{a}_U(P_m), \omega)]$, and the rest of the proof is analogous to that of Proposition 1. □

Note that the value of ambiguous persuasion varies with Receiver's preferences, as indicated by the subscript U. Consider for instance the introductory example (Sect. 2) and assume Receiver is ambiguity seeking. In this case, Receiver would then never take the safe action when ambiguity is present and hence Sender may not benefit from ambiguous persuasion. The construction of $\overline{V}_U$ would however reflect this as well so that, in this case, one would have $\overline{V}_U(p_0) = \hat{V}(p_0)$.

Furthermore, by stating that Receiver's preferences depend on the likelihood-by-likelihood posterior sets only, we do not preclude Receiver from taking into consideration only a subset of these posterior sets in order to determine his optimal action. For instance, suppose Receiver is an α-min decision maker (Gajdos et al., 2008) with utility $U(a, P_m) = \min_{p \in \phi(P_m)} E_p[u(a, \omega)]$, where $\phi(P_m) = \alpha P_m + (1-\alpha)s(P_m)$ and $s(P_m)$ is the Steiner point of P_m. The following example shows our construction of $\bar{V}$ would still work here.

Example 6. Consider the introductory example and now Receiver has the extreme case of α-min EU where $\alpha = 0$. Then $\hat{a}([\underline{p}, \overline{p}]) = \textit{brand name}$ if and only if $\underline{p} + \overline{p} \geq 2p^*$.[33] Assume $p_0 = 1/2$ and Receiver's threshold belief is $p^* = 7/8$. In this case, as illustrated by Figures 8 and 9, which depict the concave closure of v_1 and the value of ambiguous persuasion with 0-min EU receiver repsectively, the value of ambiguous persuasion is $\overline{V}(p_0) = 2/3$. This value is obtained by eliciting the posterior sets $P_{m_l} = \{0\}$ and $P_{m_h} = [3/4, 1]$. In contrast, the value of Bayesian persuasion would in this case be $\hat{V}(p_0) = 4/7 < 2/3$.

In contrast, our characterization no longer applies if Receiver's preferences at a posterior set P_m do depend on the particular message m at which it was achieved. This would be the case if Receiver updates his beliefs via the maximum likelihood rule (Gilboa and Schmeidler, 1993).

Example 7. For example, consider the following two ambiguous devices, where prior belief is $p_0(\omega_h) = 1/2$.

$\pi_1(m/\omega)$	ω_l	ω_h
m_l	1	1/3
m_h	0	2/3

$\pi_2(m/\omega)$	ω_l	ω_h
m_l	0	2/3
m_h	1	1/3

In this case, m_l is sent with probability 2/3 from device 1 and 1/3 from device 2. As a result, only the posterior resulting from device 1 is kept in the posterior set: $P_{m_l} = \{1/4\}$.

[33] Given the Steiner point of a segment is its midpoint (Gajdos et al., 2008), $s([\underline{p}, \overline{p}]) = \frac{\underline{p}+\overline{p}}{2}$.

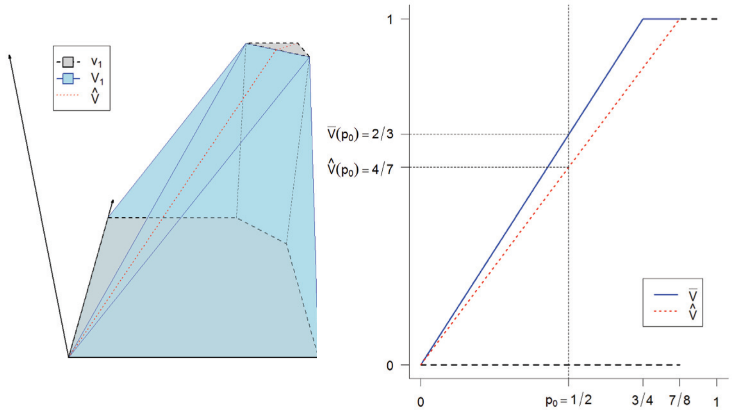

Figure 8: $V_1(P)$.

Figure 9: $\overline{V}(p)$.

For the same reason, $P_{m_h} = \{1/4\}$. Hence the maximum likelihood updated posterior sets are $P_{m_l} = P_{m_h} = \{1/4\}$. These are evidently not potentially generalized Bayes plausible.

However, the following device would lead to the same posterior set $P_{m_l} = [0, 1/4]$ under both likelihood-by-likelihood updating and maximum likelihood updating, as message m_l is sent with probability 1/2 from both devices. Similarly, $P_{m_h} = [3/4, 1]$ and the posteriors generated by maximum likelihood updating are potentially generalized Bayes plausible.

$\pi_1(m/\omega)$	ω_l	ω_h
m_l	1	0
m_h	0	1

$\pi_2(m/\omega)$	ω_l	ω_h
m_l	3/4	1/4
m_h	1/4	3/4

Although our main characterization result does not apply if Receiver uses maximum likelihood updating, Sender could still benefit from ambiguous persuasion. This includes cases where ambiguous persuasion is not beneficial with likelihood-by-likelihood updating.

For instance, consider a case when Receiver takes a safe action if $p > 1/4$ and a risky action if $p \leq 1/4$, which lead to Sender payoffs of 0 and 1, respectively. The device described above therefore has value of 1 to Sender if Receiver uses maximum likelihood updating. However, with probability-by-probability updating, our characterization result implies that Sender cannot benefit from ambiguous persuasion.

Finally, note that because we assume Sender is only interested in its *ex ante* valuation, the updating rule used by Sender is irrelevant.

7 Conclusion

In this paper, we investigate communication between an informed sender and an uninformed receiver, when the former can fully commit to his communication strategy. We show that the sender can often beneficially introduce ambiguity in his communication strategy, when agents are ambiguity averse à la Gilboa and Schmeidler (1989). The result could provide a justification for why in various situations of persuasion, the information provided by the expert is ambiguous and subject to multiple interpretations. An interesting feature of ambiguous persuasion is that the expert can make use of synonyms, duplicated messages that induce the same beliefs, and hedge himself against ambiguity. In fact, the use of synonyms (albeit in a weaker sense) turns out to be a necessary component of an optimal ambiguous persuasion device.

This key insight is robust to a valuableness requirement, consideration for more sophisticated receivers who can choose whether to listen to the expert, and a variety of ambiguity-averse preference models beyond the maxmin EU model.

We also explore its implications in two examples. In the first one, the expert can deliberately use ambiguity to muddy the waters and persuade Receiver to take a safe action, which is not ex ante optimal. In the second one, in Crawford and Sobel's (1982) "uniform-quadratic" setting, a biased expert can achieve a 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. In the latter case, an interesting observation is that, with the possibility to commit, full revelation would be optimal and "vagueness" would disappear under expected utility. Nevertheless, this is no longer the case when ambiguity is present.

A Appendix: Proofs

A.1 Proof of Lemma 1

Proof. "If". Suppose $(p_m^k)_{M\times K}$ consists of probability-by-probability updated posteriors from some K-ambiguous device $(M,(\pi_k)_K)$ with common support on M, we want to verify PGBP. By assumption, we have

$$p_m^k(\omega) = p_m^{\pi^k}(\omega) = \frac{\pi^k(m|\omega)p_0(\omega)}{\int_\Omega \pi^k(m|\omega')p_0(\omega')d\omega'}, \quad \forall m,k,\omega.$$

Also, the marginal probabilities on M by device k are $\tau^k(m) = \int_\Omega \pi^k(m|\omega')p_0(\omega')d\omega'$ for all m. By the common support property of $(\pi_k)_K$, $supp(\tau^k) = supp(\tau^j)$ for all k,j. For every device $k \in K$, we can verify that Bayes plausibility holds, i.e.,

$$\begin{aligned}\sum_{m'} \tau^k(m') p^k_{m'}(\omega) &= \sum_{m'} \left[\int_\Omega \pi(m'/\omega') p_0(\omega') d\omega'\right] p^k_{m'}(\omega) \\ &= \sum_{m'} \left[\int_\Omega \pi(m'/\omega') p_0(\omega') d\omega'\right] \frac{\pi^k(m'/\omega) p_0(\omega)}{\int_\Omega \pi^k(m'/\omega') p_0(\omega') d\omega'} = p_0(\omega),\end{aligned}$$

for all ω.

"Only if". Suppose $(p^k_m)_{M\times K}$ satisfies PGBP, with τ^k being the weights such that $\sum_m \tau^k(m) p^k_m = p_0$ for each $k \in K$. Then, we can construct K probabilistic devices such that $\pi_k(m/\omega) := \frac{\tau^k(m) p^k_m(\omega)}{p_0(\omega)}$ for each $k \in K$. This constructs an ambiguous device $(M, (\pi_k)_K)$ with common support on M, i.e., $supp(\tau^k) = supp(\tau^j)$ for all $k, j \in K$. Then, for each $k \in K$, the Bayesian posteriors of device π_k are

$$\begin{aligned}p^{\pi_k}_m(\omega) &= \frac{\pi_k(m/\omega) p_0(\omega)}{\int_{\Omega'} \pi_k(m/\omega') p_0(\omega') d\omega'} = \frac{\pi_k(m/\omega) p_0(\omega)}{\tau^k(m)} \\ &= \frac{\tau^k(m) p^k_m(\omega) p_0(\omega)}{\tau^k(m) p_0(\omega)} = p^k_m(\omega),\end{aligned}$$

for all ω. Moreover, the induced marginal distribution on M is

$$\begin{aligned}\tau^{\pi_k}(m) &= \int_\Omega \pi_k(m/\omega') p_0(\omega') d\omega' = \int_\Omega \frac{\tau^k(m) p^k_m(\omega')}{p_0(\omega')} p_0(\omega') d\omega' \\ &= \tau^k(m) \int_\Omega p^k_m(\omega') d\omega' = \tau^k(m),\end{aligned}$$

for all k, m. Hence, $(M, (\pi_k)_K)$ inherits the common support property from $(\tau^k)_K$. Hence, we find an ambiguous device $(M, (\pi_k)_K)$, satisfying the common support property, whose probability-by-probability updated vector of posteriors $(p^{\pi_k}_m)_{M\times K}$ equals to $(p^k_m)_{M\times K}$. $\square$

A.2 Proof of Lemma 2

Proof. Let $\pi'_1 = \alpha\pi_1 \oplus (1-\alpha)\pi_2$ and $\pi'_2 = (1-\alpha)\pi_2 \oplus \alpha\pi_1$. In order to avoid confusion, I denote here $\hat{a}_m(\Pi)$ the strategy of Receiver when he receives message m, which was sent through the ambiguous communication device Π. Let $\Pi = \{\pi_1, \pi_2\}$ and $\Pi' = \{\pi'_1, \pi'_2\}$.

Note that $\hat{a}_{m_1}(\Pi') = \hat{a}_{m_1}(\Pi)$ as the posterior belief at m_1 of Receiver when π'_1 is the Bayesian device is given by

$$\frac{p_0(.)\pi'_1(m_1/.)}{\pi'_1(m_1)} = \frac{p_0(.)\alpha\pi_1(m_1/.)}{\alpha\pi_1(m_1)} = \frac{p_0(.)\pi_1(m_1/.)}{\pi_1(m_1)}$$

and is therefore equal to the posterior belief at m_1 when π_1 is the Bayesian device, and posterior belief at m_1 of Receiver when π'_2 is the Bayesian device is equal to the posterior belief at m_1 when π_2 is the Bayesian device. Likewise, $\hat{a}_{b(m_1)}(\Pi') = \hat{a}_{m_1}(\Pi)$.

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

Similarly, $V(\pi_2') = (1-\alpha)V(\pi_2) + \alpha V(\pi_1)$. □

A.3 Proof of Lemma 3

Proof. First, we 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 $k \in K\backslash\{1\}$ 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, we 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 $V(\pi_1) = \max_{k \in K} V(\pi_k)$. Let $\Pi' = co((\pi_k'))_K$ be the ambiguous device that uses messages in $\cup_k M_k$.

For some value $\alpha \in [0,1]$, define π_k' as the probabilistic device such that

$$
\pi_k' := \frac{1-\alpha}{K-1}\pi_k \oplus \ldots \oplus \frac{1-\alpha}{K-1}\pi_K \oplus \ldots \oplus \alpha\pi_1 \oplus \frac{1-\alpha}{K-1}\pi_2 \oplus \ldots \oplus \frac{1-\alpha}{K-1}\pi_{k-1}.
$$

Below is this device in matrix form:

Π'	π'_1	π'_2	...	π'_K
$m_1 \in M_1$	$\alpha\pi_1(m_1/\omega)$	$\frac{1-\alpha}{K-1}\pi_2(m_1/\omega)$	...	$\frac{1-\alpha}{K-1}\pi_K(m_1/\omega)$
$b_2(m_1) \in M_2$	$\frac{1-\alpha}{K-1}\pi_2(m_1/\omega)$	$\frac{1-\alpha}{K-1}\pi_3(m_1/\omega)$	...	$\alpha\pi_1(m_1/\omega)$
...	...	...	...	...
$b_K(m_1) \in M_K$	$\frac{1-\alpha}{K-1}\pi_K(m_1/\omega)$	$\alpha\pi_1(m_1/\omega)$	...	$\frac{1-\alpha}{K-1}\pi_{K-1}(m_1/\omega)$

From lemma 2, the value of device π'_k, $V(\pi'_k)$, is $\alpha V(\pi_1) + \frac{1-\alpha}{K-1}\sum_{k\neq 1} V(\pi_k)$ for all $k \in K$. As a result, $V(\Pi') = \alpha V(\pi_1) + \frac{1-\alpha}{K-1}\sum_{k\neq 1} V(\pi_k)$.

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

For the second part, note that $V(\pi'_k) = \alpha V(\pi_1) + \frac{1-\alpha}{K-1}\sum_{k\neq 1} V(\pi_k)$ and the conclusion follows as $\alpha \to 1$.

□

A.4 Proof of Proposition 1

Proof. **If.** To prove that 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 $\tilde{P}^{-1} = co(\tilde{p}_k)_{k\geq 2} \in \underset{P^{-1}\in\Delta(\Omega)^{K-1}}{\arg\max} V_1((p_0, P^{-1}))$.

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

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

By construction, $\{p^1_m\}$ are potentially 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 generalized 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 Sender, when computed with regard to π_1 is equal to $\overline{V}(p_0)$.

Assume now that the posteriors $(\tilde{P}^{-1}_m)_M$ 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 $p_0 \in co(p^k_{m_r}, (p^k_m)_M)$; this is feasible 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 $(\tilde{P}^{-1}_m)_{M \cup \{m_r\}}$ is potentially generalized Bayes plausible where $\tilde{P}^{-1}_{m_r} = (p^k_{m_r})_{k \geq 2}$. 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 $\varepsilon\tau_1(m)$ 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 Lemma 3, 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 $\Pi = co(\pi_k)_{k \in K}$ be an ambiguous device. Let $P_m = (p^k_m)_{k \in K}$ 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 device π_1 yields the highest Sender utility, i.e., $1 \in \arg\max\limits_k(\sum_m \tau_k(m)v_k(P_m))$.

$$
\begin{aligned}
v^* &= \min_k \sum_m \tau_k(m)v_k(P_m) \\
&\leq \max_k \sum_m \tau_k(m)v_k(P_m) \\
&= \sum_m \tau_1(m)v_1(P_m) \\
&\leq \max_{P^{-1}_m} \sum_m \tau_1(m)v_1(p^1_m, P^{-1}_m) \\
&\leq \max_{P^{-1}_m} V_1(p_0, P^{-1}_m) \\
&= \overline{V}(p_0),
\end{aligned}
$$

where the last $\leq$ follows from V_1 is the concave closure of v_1 and $p_0 \in co((p^1_m)_M)$.

By Corollary 1, the value of $\overline{V}(p_0)$ is independent from K as long as $K \geq 2$. □

A.5 Proof of Lemma 4

Proof. By definition of $\overline{v}$, there exists a profile of posteriors (p^1_m) and posterior set (P^2_m) such that $p_0 = \mathbb{E}_{\tau^1} p^1_m$ and $\overline{V}(p_0) = \mathbb{E}_{\tau^1}\mathbb{E}_{p^1_m} v(\hat{a}(co(p^1_m, P^2_m)), \omega)$ by construction of $\overline{V}$.

Note that if $\mathbb{E}_{p^1_m} v(\hat{a}(co(p^1_m, P^2_m)), \omega) < \overline{v}(p^1_m)$, then it must be that $\tau^1(m) = 0$. If this were not the case, one could construct a better ambiguous device by leaving π_1 unchanged while modifying $(\pi_k)_{k\neq 1}$ so that $\hat{a}(co(p^1_m, P^2_m)) = \overline{a}(p^1_m)$, and keeping other posterior sets at other message $m' \neq m$ unchanged.[34] This new device has a strictly greater value than the original one when evaluated using π_1, and hence (by our main characterization) one can construct an ambiguous device using synonyms with this greater value. This contradicts the optimality of the original device.

Therefore at an optimal device if $\tau^1(m) > 0$, then $\mathbb{E}_{p^1_m} v(\hat{a}(co(p^1_m, P^2_m)), \omega) = \overline{v}(p^1_m)$. As a result, posterior p^1_m and also the pair $(\hat{a}(co(p^1_m, P^2_m)), p^1_m)$ are relevant.

For all m, define $h_0(p^1_m)$ as the value such that $(p^1_m, h_0(p^1_m)) \in H_0$.[35] By definition of H_0 and concavity of $\overline{V}$, it must be that $\overline{v}(p^1_m) \leq h_0(p^1_m)$ for all m.

Suppose posterior $p^1_{m'}$ is not relevant for some message m' such that $\tau^1(m') > 0$ at some optimal ambiguous device. By definition, $\overline{v}(p^1_{m'}) < h_0(p^1_{m'})$. As a result, $\mathbb{E}_{\tau^1}\overline{v}(p^1_m) < \mathbb{E}_{\tau^1} h_0(p^1_m) = h_0(p_0) = \overline{V}(p_0)$. This contradicts the optimality of the ambiguous device.

Thus at an optimal device if $\tau^1(m') > 0$, then its related posterior $p^1_{m'}$ is also relevant. □

A.6 Proof of Proposition 2

Proof. We proceed in two cases.

Consider first the case where $\overline{V}(p_0) = \overline{v}(p_0)$ and hence p_0 is relevant. In this case, under an optimal ambiguous device, an action a_0 would always be chosen by Receiver. No other action could be played in equilibrium with non-zero probability as it would not be relevant if payoffs of action a_0 were increased by an infinitesimal amount. In equilibrium, either a unique uninformative message is sent so that there are no gains to (ambiguous or even Bayesian) persuasion or there are several messages leading to the same action as our lead example — weak synonyms are necessary.

Then consider the case where $\overline{V}(p_0) > \overline{v}(p_0)$ as in example 2.

[34] In the case where this necessitates a non Bayes plausible posterior set profile, then one can add a redundant message that would never be sent from π_1 to make the posterior sets plausible again—a trick we used in the characterization proof.

[35] Even if Assumption 1 is relaxed and H_0 is not unique, one can still take the minimum of these values.

Let $\Pi = (\pi_k)_K$ be an optimal ambiguous device that uses no weak synonyms. This implies that there are no messages m and m' at which Receiver takes the same action a and for which m is sent with non-zero probability from at least one device and m' is sent with probability non-zero from at least one, potentially other, device.

Let $\mathcal{A}_k = (a_m^k, p_m^k)_{m/\tau_k(m)>0}$ be the set of action-posterior pairs that result from π_k with non-zero probability. These are necessarily relevant from lemma 4. We can show that for any $k \neq k'$, $\mathcal{A}_k = \mathcal{A}_{k'}$.

To do this, we first prove a lemma, which says that if an action is elicited with non-zero probability in one Bayesian device then it must be elicited with non-zero probability in all other Bayesian devices.

Lemma 8. *If $(a, p) \in \mathcal{A}_k$ then $(a, p') \in \mathcal{A}_{k'}$ for some p'.*

Proof of lemma 8. Suppose the statement is false for some (a, p). Denote by m the message at which action a is chosen by Receiver when π_k is used. Hence action a is played with positive probability under Bayesian device π_k but with zero probability under device $\pi_{k'}$.

We perturbe the game by decreasing sender's payoff to the action a by an arbitrarily small amount, that is,

$$\begin{aligned} v^\varepsilon(a, \cdot) &= v(a, \cdot) - \varepsilon \\ v^\varepsilon(a', \cdot) &= v(a', \cdot) \text{ for } a' \neq a. \end{aligned}$$

for small $\varepsilon > 0$. In this case, the value $\overline{V}(p_0)$ should not change after the perturbation. To see this, the value of device Π when evaluate by device $\pi_{k'}$, $V(\pi_{k'})$, would not change since action a was played with probability zero under device $\pi_{k'}$. Furthermore, by optimality of the ambiguous device, this value of device Π when evaluated by device $\pi_{k'}$ must always equal to $\overline{V}(p_0)$ before and after the perturbation. As a result, we can show that the decrease in payoffs at action a does not modify the hyperplane H_0.

Denoting h_0^ε the value function along H_0 in the perturbed game, we have that $h_0^\varepsilon(p_0) = h_0(p_0)$ since H_0 remains unchanged. Let $M(a)$ be the set of messages such that the posteriors under device k namely p_m^k form a relevant action-posterior pair (a, p_m^k) with a.[36] Thus the value at all other action-posterior pairs used by π_k with actions differ from a have not decreased either. That is, for all $m' \notin M(a)$, we have $h_0^\varepsilon(p_{m'}^k) = h_0(p_{m'}^k)$.

Then for device k, the fact that H_0 is an hyperplane and $p_0 = \sum_{m\in M} \tau(m) p_m^k$ in device k

[36]Note that for Lemma 8 we don't need the device to be straightforward.

imply

$$\begin{aligned} h_0(p_0) &= \sum_{m \in M(a)} \tau(m) h_0(p_m^k) + \sum_{m' \notin M(a)} \tau(m') h_0(p_{m'}^k) \\ h_0^\varepsilon(p_0) &= \sum_{m \in M(a)} \tau(m) h_0^\varepsilon(p_m^k) + \sum_{m' \notin M(a)} \tau(m') h_0^\varepsilon(p_{m'}^k) \end{aligned}$$

By Assumption 2, all the relevant action posterior pairs should remain relevant so for all $m \in M(a)$,

$$h_0^\varepsilon(p_m^k) = \overline{v}^\varepsilon(p_m^k) = \overline{v}(p_m^k) - \varepsilon.$$

This contradicts $h_0(p_0) = h_0^\varepsilon(p_0)$ and finishes the proof the the Lemma 8.

Lemma 9. *$\mathcal{A}_k = \mathcal{A}_{k'}$ for all k and k'.*

Proof of Lemma 9. Suppose not, then there is some $(a,p) \in \mathcal{A}_k$ and $(a,p) \notin \mathcal{A}_{k'}$ for some k'. Then, from Lemma 8, there must be some $p' \neq p$ such that $(a,p') \in \mathcal{A}_{k'}$. Therefore both (a,p) and (a,p') are relevant.

Without loss of generality, assume $p'(\omega) < p(\omega)$.[37] Consider the following perturbation:

$$\begin{aligned} v^\varepsilon(a,\omega) &= v(a,\omega) - \varepsilon \\ v^\varepsilon(a,\omega') &= v(a,\omega') + \frac{p'(\omega)}{1-p'(\omega)}\varepsilon \text{ for } \omega' \neq \omega \\ v^\varepsilon(a',\cdot) &= v(a',\cdot) \text{ for } a' \neq a. \end{aligned}$$

It decreases Sender's payoff of action a at ω by an arbitrarily small ε and increase it in other states by $\frac{p'(\omega)}{1-p'(\omega)}\varepsilon$ such that $\mathbb{E}_{p'} v^\varepsilon(a,\omega)$ is left unchanged. If Assumption 2 is satisfied and ε is small enough, then (a,p') and (a,p) would still be relevant in this perturbed game.

Yet, by the same reasoning as before the value of the ambiguous device when evaluated with $\pi_{k'}$ as well as the optimal value of Π will not change so that $h_0^\varepsilon(p_0) = h_0(p_0)$. And as before, we have that $h_0^\varepsilon(p_{m'}^k) = h_0(p_{m'}^k)$ for all $p_{m'}^k \neq p$ such that $(a_k, p_{m'}^k) \in \mathcal{A}_k$ for some $a_k \neq a$. Again this relies on the fact that the ambiguous device uses no weak synonyms so no other action-posterior pair in $\mathcal{A}_k$ would be modified in this perturbation. We therefore have that $h_0^\varepsilon(p) = h_0(p)$. Yet, by construction, we have that $\mathbb{E}_p v^\varepsilon(a,\omega) < \mathbb{E}_p v(a,\omega) = \overline{v}(p) = h_0(p)$. Thus, (a,p) is not relevant in the perturbed game, in contradiction with Assumption 2. This ends the proof of Lemma 9.

Because we assumed that Π did not use weak synonyms, it must be that every relevant action posterior pair (a,p) must be induced from a distinct message. Hence the lemma above implies that at each message (that is sent with non-zero probability by any π_k) the posterior set P_m is a singleton.

[37] In this proof we assume without loss that posteriors have finite support on Ω. If Ω is a continuum and p and p' are atomless, one can find some measurable event $B \subseteq \Omega$ where $p'(B) < p(B)$ and the same argument applies.

As a result, Π elicits exactly the same actions as some Bayesian device π. There is therefore no gains to ambiguous persuasion (over that of Bayesian persuasion). $\square$

A.7 Proof of Lemma 5

Proof. Given the set of posteriors $\{I_A, I_B\}$, if Receiver takes an action $x + b$, Receiver's maxmin EU can be written as (with a slight abuse of notation)

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

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

$$2x\mathbb{E}_{[0,c]}\omega - \mathbb{E}_{[0,c]}\omega^2 \quad \leq \quad 2x\mathbb{E}_{[c,1]}\omega - \mathbb{E}_{[c,1]}\omega^2, \tag{4}$$

which simplifies into

$$x \quad \geq \quad \frac{l+c}{3} \equiv x^*. \tag{5}$$

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 $\mathbb{E}u(x+b) = h_2(x)$ for $x \leq x^*$ and $\mathbb{E}u(x+b) = h_1(x)$ 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 unique optimal action when observing m_A is

$$x^* + b \quad = \quad \frac{l+c}{3} + b. \tag{6}$$

Hence $\hat{a} = a^* = \frac{l+c}{3} + b$.

Sender's ex-ante utility is

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

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

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

Combined with the domain restriction that $c \in [0, l)$, we obtain the desired result. □

A.8 Proof of Lemma 6

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

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

Thus,

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

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 $l' \in [0, 1]$

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

and

$$\tilde{V}(b) + \tilde{V}(l + l' - b) \geq \tilde{V}(l) + \tilde{V}(l') \qquad \text{if } l + l' \geq 2b. \tag{8}$$

To prove inequality (7), 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 (8), 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 $l \in [0,b]$, and

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

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

A.9 Proof of Proposition 4

Proof. By symmetry, and following the calculation preceding Lemma 6, we have that Receiver's optimal action upon observing either message $m_{i,A}$ or message $m_{i,B}$ is

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

Note that Sender's expected payoff can be written

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

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

$$\hat{a}_1 = a_1^* \equiv x_1^* + b = \frac{1}{3n} + b,$$

and Sender's expected payoff is

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

which is maximized when

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

or

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

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 ambiguous communication 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 $\hat{a}_1 = x_1^* + b = 1/(3n) + b$.

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 (9), 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.10 Discussion of KLQ

KLQ consider cheap-talk communication, where Sender cannot commit to a communication rule. Therefore, 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.

Thus, according to our (and KLQ's) characterization of Receiver's best response in (6), it must be that

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

or

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

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 (6) and Sender's incentive compatibility constraint. Let us focus on two adjacent intervals of length l and l', we must have

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

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.

A.11 Proof of Corollary 2

Proof. Dynamic Consistency $\Leftarrow$ Rectangularity. See 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. $\square$

A.12 Proof of Proposition 6

Proof. Suppose that $R = \bar{R}$ and there exists a potentially Bayes plausible selection $(q^*_m)_M \in (Q^*_m)_M$ of the supporting posteriors. By potential Bayes plausibility, there exists $\bar{\tau} \in \Delta M$ such that $\sum_m \bar{\tau}(m) q^*_m = p_0$. By definition of a consistency closure, $(\bar{\tau}, \mathbf{q}^*) \in co(\bar{R})$.

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

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

A.13 Proof for the example of Section 6.3

Proposition 9. *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 10 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. We 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.

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

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.

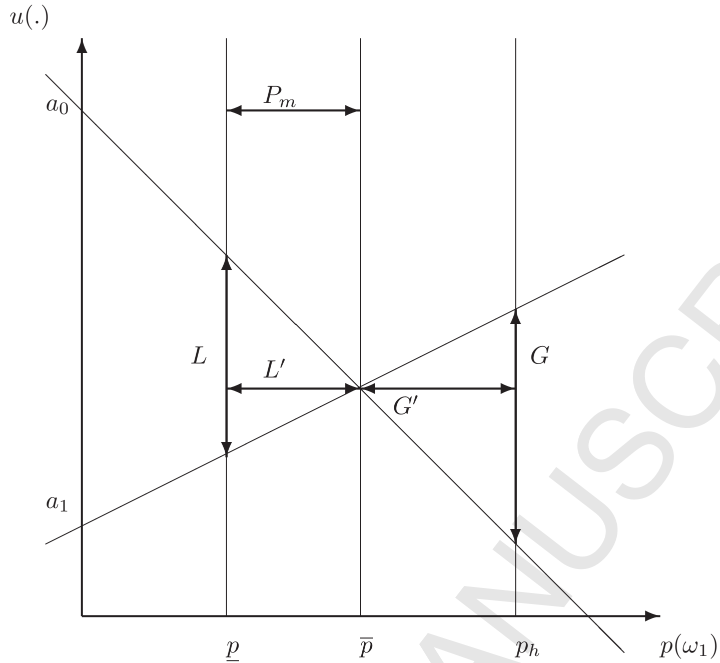

Figure 10: Gains and Losses from Embedded Information

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

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

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[38]

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

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 two new functions:

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

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

Given $\underline{\pi}(h)G(a,b) - \underline{\pi}(m)L(a,b) < 0 \Leftrightarrow \frac{G(a,b)}{L(a,b)} < \frac{\underline{\pi}(m)}{\underline{\pi}(h)}$, 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[39].

- Thus Π valuable implies condition (1) in the proposition. The exact same argument is used to show that Π valuable implies condition (2).

[38]
$$\begin{aligned} L(a,b,c) &= (1-\underline{p})a - \tfrac{a}{a+c-b}.[\underline{p}c + (1-\underline{p})b] - (1 - \tfrac{a}{a+c-b})(1-\underline{p})a \\ &= \tfrac{a}{a+c-b}[(1-\underline{p})(a+c-b) - \underline{p}c - (1-\underline{p})b - (1-\underline{p})(a+c-b) + (1-\underline{p})a] \\ &= \sigma(a,b,c)[-\underline{p}c - (1-\underline{p})b + (1-\underline{p})a] \\ &= \sigma(a,b,c)[a - b - \underline{p}(a+c-b)] \end{aligned}$$

[39] Pick a and b randomly and $c = \frac{(1-\overline{p})(a-b)}{\overline{p}}$.

- 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. □

A.14 Proof of Proposition 8

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 $p^{-1}(a_0)$, $u(a_0, p_m) > u(\hat{a}(P_m), p_m)$.

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

$$\begin{aligned} U(\Pi) &= \min_k \sum_m \tau_m^k u(\hat{a}(P_m), p_m^k) \\ &\geq \min_k \sum_m \tau_m^k u(a_0, p_m^k) \\ &\geq \min_k \, u(a_0, p_0) \geq U_0 \quad \square \end{aligned}$$

A.15 Proof of Corollary 3

Proof. Proposition 1 shows there exists a sequence of devices such that their MaxMin value (for a sender) converges to $\overline{V}(p_0)$. Note that this MaxMin value is, by definition, equal to $\min_{v \in \mathbf{v}} v$ as defined above. As a result, given Sender's utility is greater than $\min_{v \in \mathbf{v}} v$ for each device in the sequence, then the value of ambiguous persuasion is greater or equal to $\overline{V}(p_0)$.

Furthermore, Lemma 3 showed that for any given device of MaxMax value $\max_{v \in \mathbf{v}} v$, there exists a sequence of devices such that their MaxMin value converge to $\max_{v \in \mathbf{v}} v$. This implies that the value of ambiguous persuasion to an ambiguity seeking sender cannot be greater than to an ambiguity averse sender. As a result, $\overline{V}(p_0)$ is an upper bound on the value of $\max_{v \in \mathbf{v}} v$ and therefore an upper bound on the value of any sender whose preferences satisfy betweenness. □

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 (2017). Hard evidence and ambiguity aversion. *Theory and Decision 82*(3), pp. 327–339.

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

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

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). 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 (2016). The design of ambiguous mechanisms. *The Review of Economic Studies 84*(1), 237–276.

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.

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

Frankel, A. (2014). 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*, pp. 179–242. Cambridge University Press.

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

Gilboa, I. and D. Schmeidler (1993). Updating ambiguous beliefs. *Journal of Economic Theory 59*(1), 33–49.

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.

Jaffray, J. Y. and F. Philippe (1997). On the existence of subjective upper and lower probabilities. *Mathematics of Operations Research 22*, 165–185.

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

Kellner, C. and M. T. Le Quement (2017). Modes of ambiguous communication. *Games and Economic Behavior 104*, 271–292.

Kellner, C. and M. T. Le Quement (2018). Endogenous ambiguity in cheap talk. *Journal of Economic Theory 173*, 1 – 17.

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 (2017). Disclosure to a psychological audience. *American Economic Journal: Microeconomics,* forthcoming.

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

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

Maccheroni, F., M. Marinacci, and A. Rustichini (2006). Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica 74*(6), 1447–1498.

Mukerji, S. and J.-M. Tallon (2004). An overview of economic applications of david schmeidler's models of decision making under uncertainty. *Uncertainty in economic theory 13*, 283.

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. Working paper presented at ESSET 2015.

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