

## States

A state is defined as a pair of a joint probability table and a Bayesian (Causal) Network (BN). The state prior first samples a BN (uniformly distributed) and then samples a probability table based on the BN. In the base model, there are three possible BNs: One for the case that A depends on C, one for the case that the reverse is true, i.e. C depends on A and one for the case that A and C are independent. For building the joint probability tables, conditional or marginal probabilities are sampled from a uniform distribution over the interval from 0 to 1 as shown in Table 1 .

Bayes Net	sampled from uniform(0,1)
<i>A implies C</i>	$P(C A), P(C \neg A), P(A)$
<i>C implies A</i>	$P(A C), P(A \neg C), P(C)$
<i>A independent C</i>	$P(A), P(C)$

Table 1: Probabilities that are sampled from a uniform distribution over the interval [0,1] for the three different Bayes Nets.

## Meaning Function

The literal meaning of an utterance is determined by its corresponding probability: it must exceed a given threshold  $t$  for making the utterance be true. In the first base model, the speaker can choose among four utterances,  $A$ ,  $C$ , *If A, then C* and *If C, then A*.  $P(A)$  and  $P(C)$  correspond to the simple utterances  $A$  and  $C$  and are just the sums of the left column and the first row respectively while  $P(C|A)$  and  $P(A|C)$  correspond to the conditional utterances *If A, then C* and *If C, then A*. Since  $P(C, A) = P(C|A) * P(A) = P(A|C) * P(C)$ , the conditional probabilities can be written as shown in Equation 1a and 1b .

$$P(C|A) = \frac{P(C, A)}{P(A)} \quad (1a)$$

$$P(A|C) = \frac{P(C, A)}{P(C)} \quad (1b)$$

The resulting condition for the meaning function can be formulated in two manners as can be seen in Figure 1 for the utterance *If A, then C*. The first is simply the direct implementation of equation 1a being greater than the threshold. The reasoning behind the second expression is different: we want the probability of  $A$  to be greater to a certain extent than the probability of  $\neg A$  given  $C$ .  $P(A|C)$  and  $P(\neg A|C)$  are obtained by normalizing the upper row of Table 2 yielding  $\frac{P(A,C)}{z}$  and  $\frac{P(\neg A,C)}{z}$ . We need  $P(A|C) = \frac{P(A,C)}{z}$  to be greater than the threshold  $t$ , therefore  $\frac{P(A,C)}{z} > t$  and since  $\frac{P(A,C)}{z} + \frac{P(\neg A,C)}{z} = 1$ , the following must hold:  $\frac{P(\neg A,C)}{z} < 1 - t$ . Summarized, we get  $\frac{P(A,C)}{t} > z$  and  $\frac{P(\neg A,C)}{1-t} < z$  which results in  $\frac{P(A,C)}{t} - \frac{P(\neg A,C)}{1-t} > 0$  when combined. This last expression is equivalent to the second condition in Figure 1 .

```

1 // utterance == "If C, then A"
2 state['jointP'][0] > (state['jointP'][1] + state['jointP'][0]) * threshold // (1)
3
4 (1-threshold) * state['jointP'][0] > state['jointP'][1] * threshold // (2)

```

Figure 1: Condition for the utterance *If C, then A* to be true expressed in two different ways.

	A	$\neg A$
C	$P(C, A)$	$P(C, \neg A)$
$\neg C$	$P(\neg C, A)$	$P(\neg C, \neg A)$

Table 2: Joint Probability Table.

In order to get discrete values, we define a discrete version of the Beta distribution as shown in Figure 2 . It is important to set the granularity of the bins not too broad. When it was set to five, errors due to rounding had a huge impact which was manifested by different results for the two (equal) expressions from Figure 1 . Setting

```

1 var granularity = 20
2 var roundTo3 = cache(function(x){
3   return Math.round(x*1000)/1000
4 })
5 )
6
7 var midBins = map(function(x) {roundTo3(x/granularity + 1/(2*granularity))},
8   _..range(0,granularity))
9
10 var DiscreteBeta = cache(function(a, b){
11   Infer({model: function(){
12     categorical({
13       vs:midBins,
14       ps:map(function(x){
15         Math.exp(Beta({a, b}).score(x))
16       }, midBins)
17     })
18   })
19 })

```

Figure 2: WebPPL Code for the discrete version of the beta distribution.

the granularity value of 10 was sufficient to get rid of this problem. Also, the results for the distribution of the pragmatic speaker were, independent from the formulation of the meaning function, substantially different depending on the granularity value. This is demonstrated in Figure 3 .

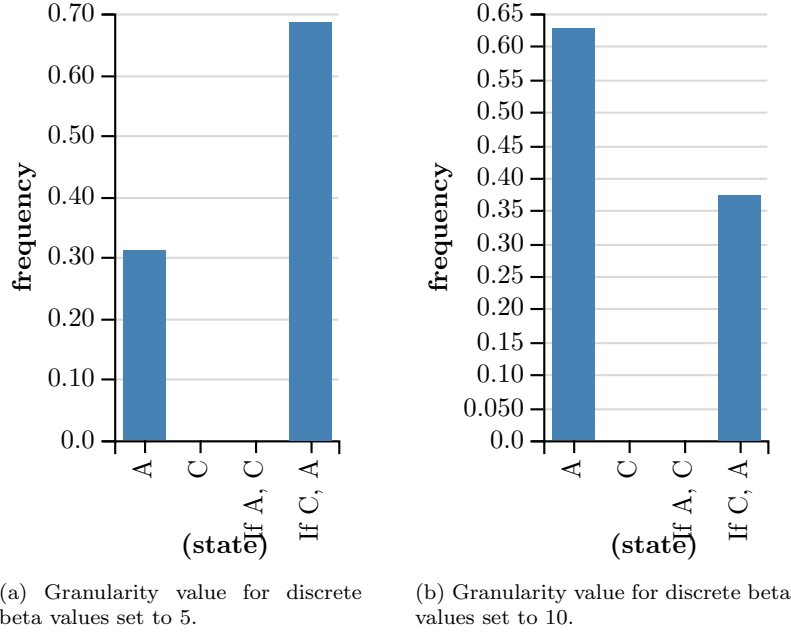


Figure 3: Speaker distribution over utterances for different granularity values of the discrete beta distribution from which we sample to get the prior joint probability tables of the listener. The pragmatic constant  $\alpha$  was set to 5 and the speaker's joint probability table was  $[0.2, 0.0, 0.8, 0.0]$  ( $P(C, A), P(C, \neg A), P(\neg C, A), P(\neg C, \neg A)$ ).

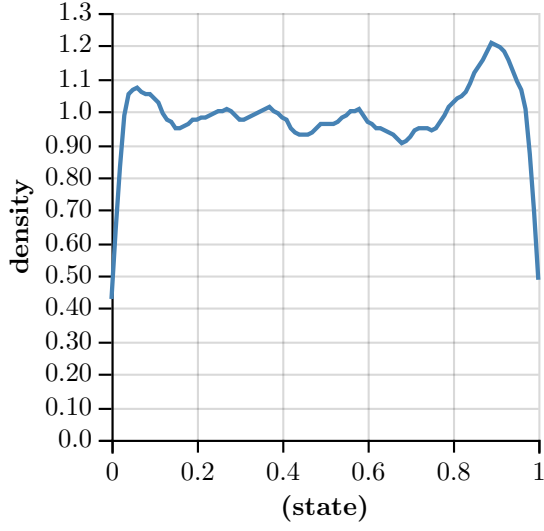
## Literal Listener

The Literal Listener is at the end of the reasoning chain in the rational speech act model. He conditions the statePrior on the meaning of the utterance that he heard. The prior distributions for  $P(A|C)$ ,  $P(A)$  and  $P(\text{BayesNet})$  from the literal listener, i.e. he has not heard any utterance yet, are shown in Figure 19. When the Literal Listener hears an utterance, he only considers those probability tables where the corresponding probability of the utterance exceeds a given threshold. Figure 4 and Figure 5 show the probability distributions for  $P(C)$  and  $P(C|A)$  of the literal Listener for all possible utterances that he may have heard. Figure 6 shows the WebPPL code for the Literal Listener.

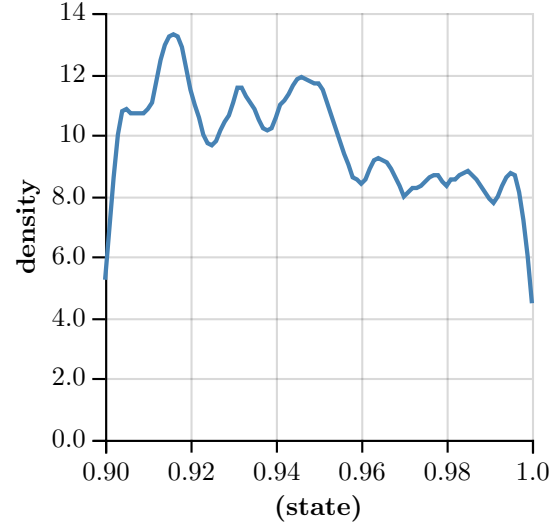
In Figure 4, when the literal listener hears  $C$ , the distribution of  $P(C)$  is the equally weighted sum of the normalized pdfs of  $C$  for the three different Bayes Networks where  $C > 0.9$ . In the two Bayes Nets  $C \text{ implies } A$  and  $A \text{ independent } C$ ,  $C$  is uniformly distributed (from 0.9 to 1, the pdf is:  $1/(1 - 0.9) = 10$ ) and in the Bayes Net  $A \text{ implies } C$ , the distribution looks like a broad bell as shown in Figure 19a. Therefore  $C$  lies in the interval  $[0.9, 1]$  with slightly decreasing pdf values for increasing  $C$  values.

Figure 8 shows the Literal Listener's distributions over the three Bayes Networks. These distributions result from conditioning on the corresponding probabilities of the heard utterances for each Bayes Network respectively. Figure 7 shows the distributions for  $P(C|A)$ , i.e. the Literal Listener heard the utterance *If A, then C*, for each of the three Bayes Nets. In the case of the two uniform distributions,  $P(P(C|A) > 0.9) = 0.1$  while  $P(P(C|A) > 0.9) \approx 0.16$  for the network  $C \rightarrow A$ . Therefore, the Literal Listener puts equal probability mass on the Bayes Nets  $A \rightarrow C$  and  $A \text{ independent } C$  and most on the network  $C \rightarrow A$ .

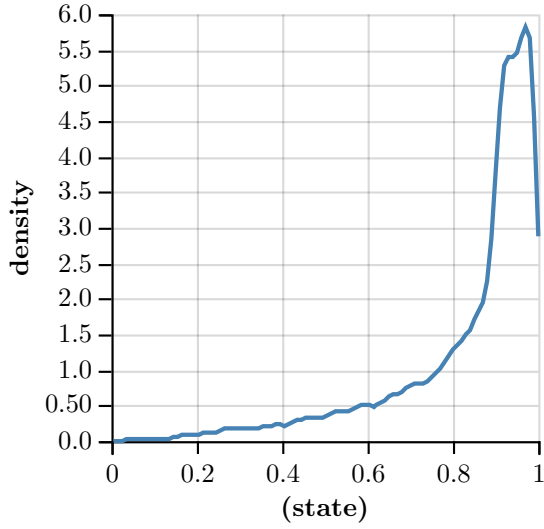
die anderen nachvollziehen!!



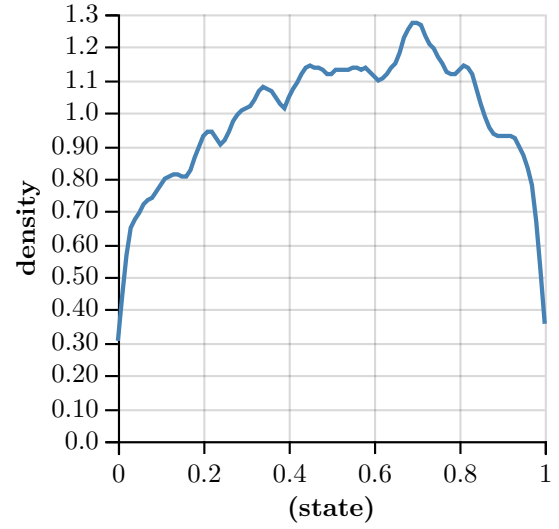
(a) LL hears *A*.



(b) LL hears *C*.

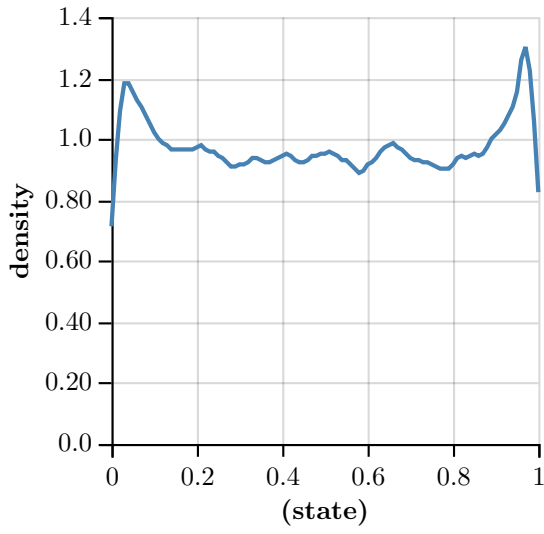


(c) LL hears *If A, C*.

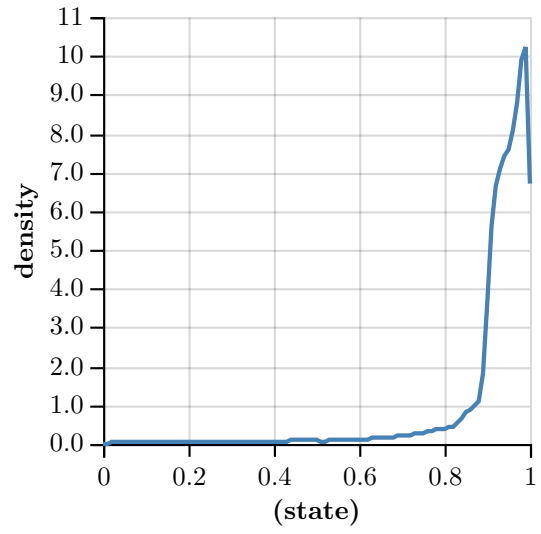


(d) LL hears *If C, A*.

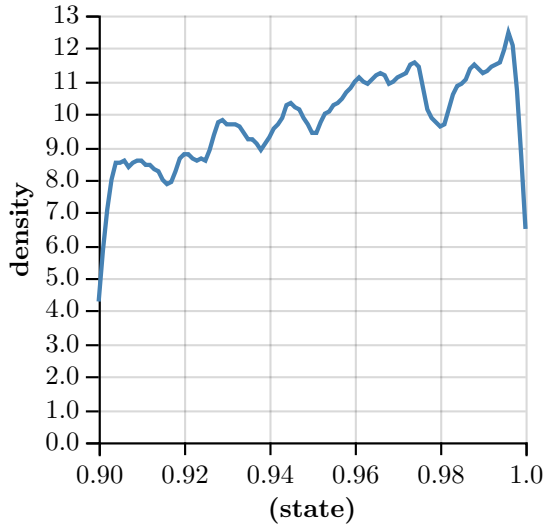
Figure 4: Distributions  $P(C)$  the Literal Listener infers depending on the heard utterances when the threshold of the meaning function is set to 0.9.



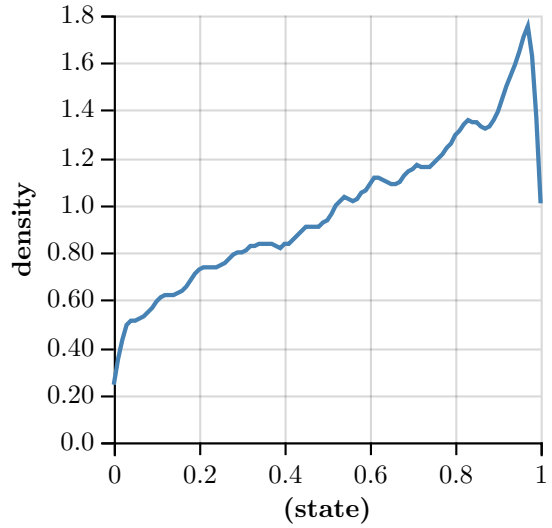
(a) LL hears *A*.



(b) LL hears *C*.



(c) LL hears *If A, C*.



(d) LL hears *If C, A*.

Figure 5: Distributions  $P(C|A)$  the Literal Listener infers depending on the heard utterances when the threshold of the meaning function is set to 0.9.

```

1  var literalListener = function(utterance, focus){
2    Infer({method: "enumerate",
3      model: function(){
4        var state = sample(statePrior());
5        condition(meaning(utterance, state))
6        if(focus == 'BN'){
7          return state['causalNet']
8        }
9        else{return state}
10      }}}
11 }

```

Figure 6: WebPPL code for the Literal Listener.

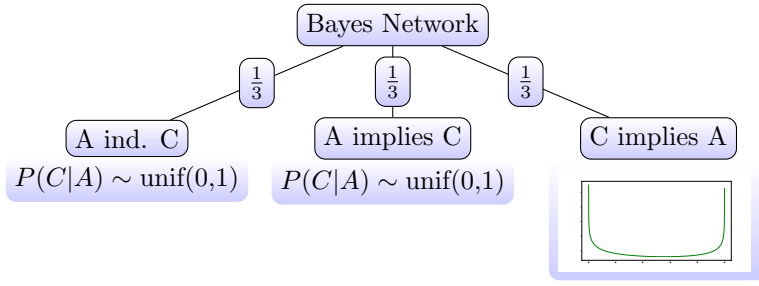


Figure 7: Distributions of  $P(C|A)$  for each Bayes Network.

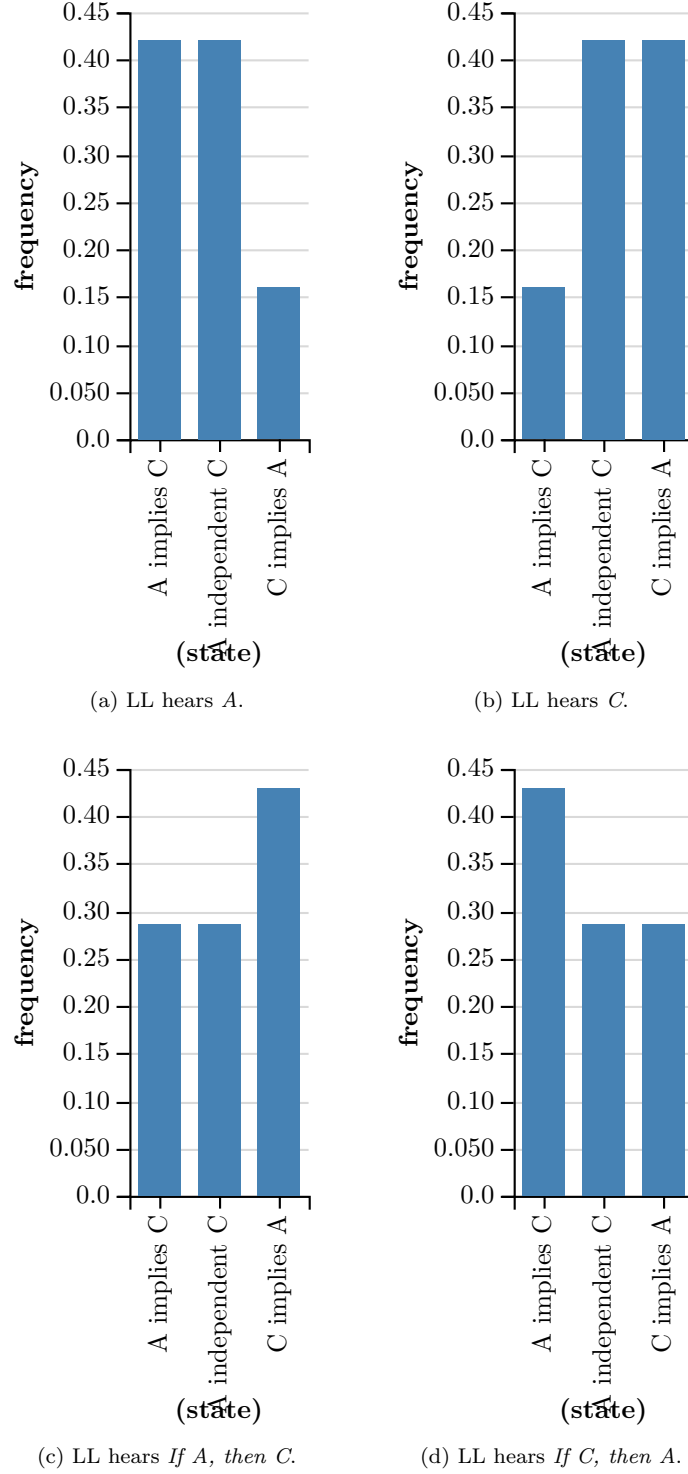


Figure 8: Literal Listener's distributions over Bayes Networks.

## Pragmatic Speaker

There are two versions of the pragmatic speaker. In the first, the pragmatic speaker is implemented as minimizing the difference between the joint probability table she has in mind and the joint probability table of the literal listener. This difference is represented by the *Kullback-Leibler divergence* ( $KL_D$ ) which measures how different two probability distributions are, i.e. how well a distribution  $Q$  approximates a true distribution  $P$ . The greater  $KL_D$  is, the greater is the difference of the two distributions. It is always positive and for identical distributions it is equal to 0. The factor statement in the code of the Pragmatic Speaker (Figure 10) adds the argument value to the log probability of the current execution. The greater  $KL_D$  is, the smaller the log probability shall be and therefore we use  $-KL_D$  inside the factor statement.

In the second version of the pragmatic speaker, the Literal Listener does not return the entire state, but the marginalized distribution for the three Bayes Networks. The pragmatic speaker has a Bayes Network in mind and chooses an utterance proportional to the probability that the Literal Listener puts on that specific Bayes Network.

$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \quad (2)$$

Figure 9: Definition of the Kullback-Leibler divergence for discrete probability distributions  $P$  and  $Q$ . It is the amount of information that is lost when  $P$  is approximated by  $Q$ . It is always  $\geq 0$ . The best value 0 is obtained if  $P$  and  $Q$  are identical. When  $Q$  overestimates an event, the quotient of the log is  $< 1$  and therefore this summand is negative which brings the result of  $D_{KL}$  closer to 0. The worst effect on the Kullback-Leibler divergence is when an event is underestimated to a great extent since then the logarithm becomes large.

```
1 var speaker = cache(function(state, goal) {
2   return Infer({model: function(){
3     var utterance = uniformDraw(utterances)
4     var LL = literalListener(utterance, goal)
5     if(goal == 'probs'){
6       // speaker wants to transmit the probabilities
7       var LLprob = sample(LL)['jointP']
8       var KLdiv = -KL(state, LLprob)
9       factor(alpha * (KLdiv - cost(utterance)))
10    }else{
11      // speaker wants to transmit the causal net
12      factor(alpha * LL.score(state))
13    }
14    return utterance
15  })
16 })
```

Figure 10: WebPPL code for the pragmatic speaker.

## Speaker Results

### Version 1: Transferring Probabilities

For all plots, the granularity of the beta distribution was set to 20 and the value for the optimality parameter  $\alpha$  was set to 5.

The results for the speaker when her state probability table is set to the one in Table 3a which are shown in Figure 11a, seem reasonable. Since  $P(A) = 1$ , it is a good choice for the speaker to produce the utterance  $A$ . It is also reasonable to choose the utterance *If C, then A* for the given state probabilities since  $P(A|C)$  is also 1. The remaining two utterances also receive some probability mass which is as expected since  $P(C|A)$  and  $P(A)$  are both high (0.8).

When considering the state probabilities of Table 3b the speaker's distribution is similar as shown in Figure 11b. The most likely utterance is  $A$  ( $> 0.6$ ) followed by *If C, then A*. The speaker does not put any probability on the utterances  $C$  and *If A, then C* whose corresponding probabilities are both 0.2. However, when the speaker is less pragmatic, i.e. the optimality parameter  $\alpha$  is smaller (here  $\alpha = 5$ ), there is some probability mass also on these two utterances.

	A	$\neg A$
C	0.8	0
$\neg C$	0.2	0

probs	values
$P(A C)$	1
$P(C A)$	0.8

(a) speaker probabilities Ex. 1.

	A	$\neg A$
C	0.2	0
$\neg C$	0.8	0

probs	values
$P(A C)$	1
$P(C A)$	0.2

(b) speaker probabilities Ex. 2.

Table 3: Exemplary speaker probability tables.

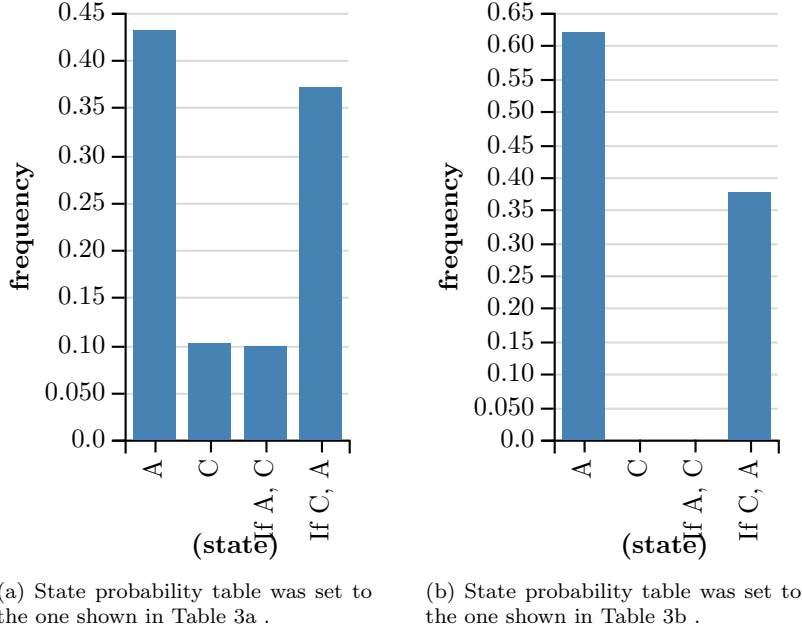


Figure 11: Probability distributions over utterances for the pragmatic speaker given the state probabilities of Table 3.

When the speaker's state probability is only distributed on those events where  $\neg A$  holds (see Table 4), and therefore  $P(A) = 0$ , most of the possible utterances are not reasonable for the speaker. When considering the state probabilities of Table 4a, the best would be to say that  $A$  is false since  $P(\neg A) = 1$  or that  $A$  is false when  $C$  is true since  $P(\neg A|C)$  is also 1. The only possible utterance which is not completely false is  $C$  since  $P(C) = 0.8$ . This is also the utterance that the speaker chooses with highest probability (0.55). The remaining probability mass is put on the utterance *If A, then C* which is not true considering the speaker's state probability. In section , the distributions over the entries of the Literal Listener's joint probability tables are plotted. With the speaker's state probabilities from Table 4a, the entries on the left column, i.e. where  $A$  is true, are always overestimated by the Literal Listener. The worst for the Kullback-Leibler divergence, however, is the underestimation of events. When the Literal Listener hears  $A$  or *If C, then A*, the entry  $\neg A, C$  is highly underestimated ( $< 0.1$ ), see Figures 20a and 23a. For the other two utterances, there is no such high underestimation and therefore the pragmatic speaker distributes the probability mass on these utterances ( $C$  and *If A, then C*).

When these state probabilities are switched, see Table 4b, the probability mass is distributed among the utterances *If A then C* ( $> 0.9$ ) and *If C, then A* which are both not true. For the simple utterances  $A$  and  $C$ , the underestimation for the entry  $\neg A, \neg C$  is very large since the state probability is 0.8 and the probability is distributed on values only up to 0.08. For the conditional utterances, there is also an underestimation, which is however less distinctive. Therefore the speaker prefers the conditional utterances and among these, the

$$P_S(u|w) \propto \exp(\alpha U(u; w)) \quad (3)$$

Figure 12: Definition Pragmatic Speaker. She is represented as a distribution over possible utterances.



	A	$\neg A$
C	0	0.8
$\neg C$	0	0.2

	A	$\neg A$
C	0	0.2
$\neg C$	0	0.8

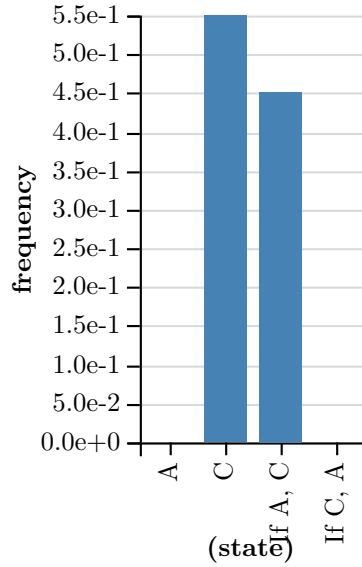
probs	values
$P(A C)$	0
$P(C A)$	0
$P(\neg A C)$	1
$P(C \neg A)$	0.8

(a) speaker probabilities  
Ex. 3.

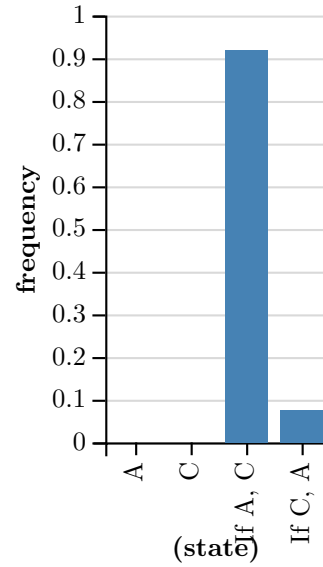
probs	values
$P(A C)$	0
$P(C A)$	0
$P(\neg A C)$	1
$P(C \neg A)$	0.2

(b) speaker probabilities  
Ex. 4.

Table 4: Exemplary speaker probability tables.



(a) State probability table was set to the one shown in Table 4a .



(b) State probability table was set to the one shown in Table 4b .

Figure 13: Probability distributions over utterances for the pragmatic speaker given the state probabilities of Table 4 .

utterance *If A, then C* where the entry for  $C, \neg A$  is rather overestimated and not underestimated as it is the case for the utterance *If C, then A*.

## Version 2: Transferring Causal Networks

The results for the speaker's probability distributions over utterances, given that she wants to transmit the underlying causal network, are shown in Figure 14 . They can be understood considering the Literal Listener's distributions over Bayes Networks as shown in Figure 8 . As desired, when the speaker wants to transfer that event  $A$  is independent of event  $B$ , she chooses the utterance  $A$  or  $C$  with equal probability ( $> 0.4$ ). However, in order to transfer the conditional networks ( $A \text{ implies } C, C \text{ implies } A$ ), in most cases the speaker chooses the contrary conditional utterance, i.e. *If C, A* and *If A, C*, or the antecedent ( $A, C$ ).

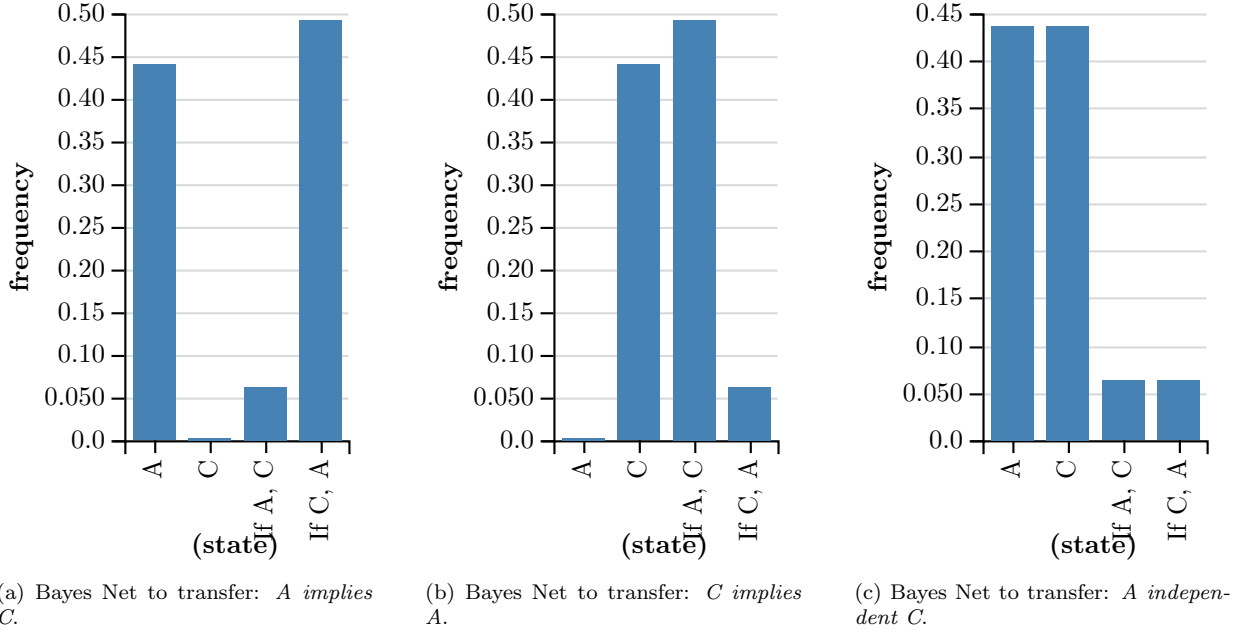


Figure 14: Speaker's probabilities for utterances given that she wants to transmit the underlying causal relations.

## Pragmatic Listener

$$P_L(w|u) \propto P_S(u|w)P(w) \quad (4)$$

Figure 15: Definition of Pragmatic Listener. He is represented as a distribution over worlds (states).

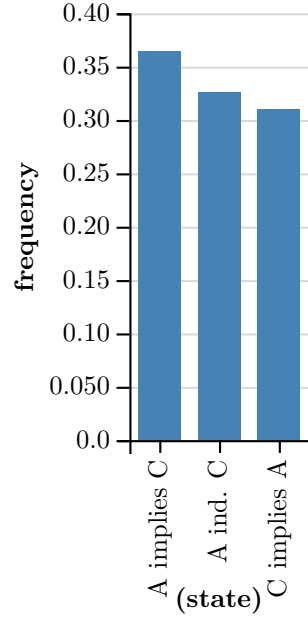
The pragmatic listener is defined as shown in Equation 15 and as WebPPL code in Figure 16. The distributions of the Pragmatic Listener over states, that he infers given that the Literal Listener is represented as the marginal distribution over joint probabilities, are shown in Figure 17. The pragmatic listener puts more probability on the state that  $A \implies C$  than on the state where  $A$  and  $C$  are independent, as is desirable. Yet, the difference is only marginal, the probability mass is almost uniformly distributed over the three causal networks. The results look very different when the speaker aims at transferring the underlying causal network instead of the joint probabilities as can be seen in Figure 18. The Pragmatic Listener interpretes the simple utterances, e.g.  $A$ , as the corresponding event implying the other event ( $A \implies C$ ) or as two independent events. For the conditional utterances, most of the probability mass is put on the contrary implication, e.g. on  $C \implies A$  for the utterance *If A, C*, as before, but the remaining two causal network states only receive very little probability mass (both 0.1).

```

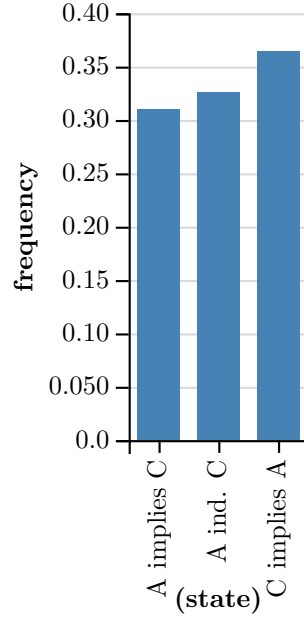
1 var listener = function(utterance, goal){
2   Infer({model: function(){
3     var state = sample(statePrior());
4     observe(speaker(state[goal], goal), utterance)
5     return state['causalNet']
6   }})
7 }

```

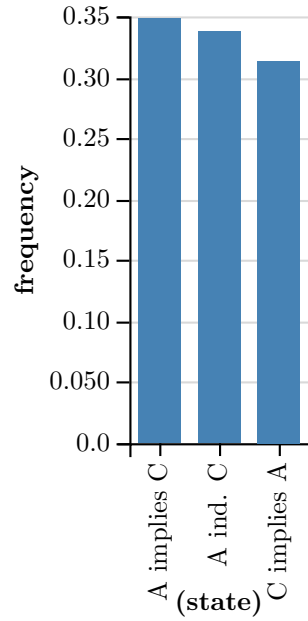
Figure 16: WebPPL code for the Pragmatic Listener.



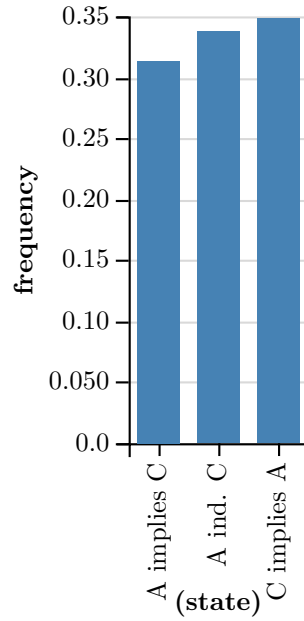
(a) Heard utterance: *A*.



(b) Heard utterance: *C*.

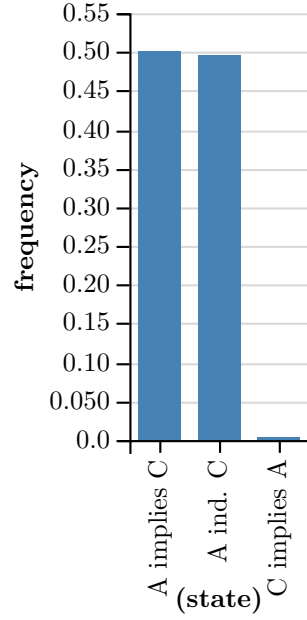


(c) Heard utterance: *If A, C*.

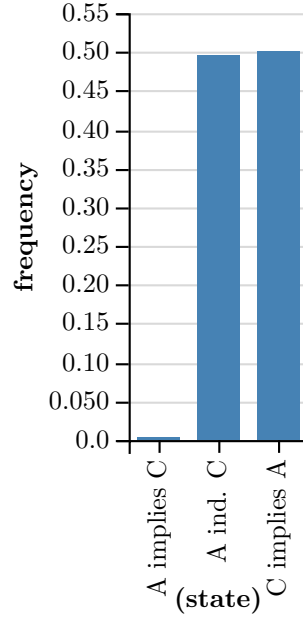


(d) Heard utterance: *If C, A*.

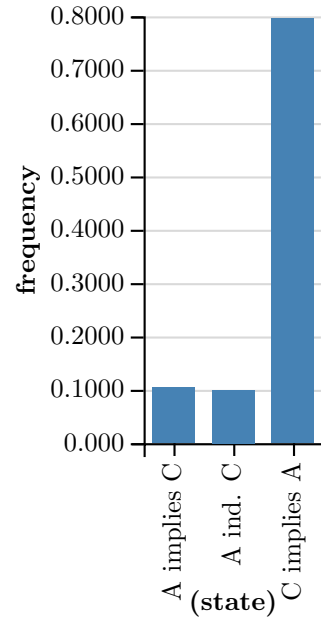
Figure 17: Results for the Pragmatic Listener depending on the utterances he has heard given that the Literal Listener is represented as distribution over joint probabilities. Due to the time consumption for the computations, the granularity for the discrete beta distribution was set to 10.



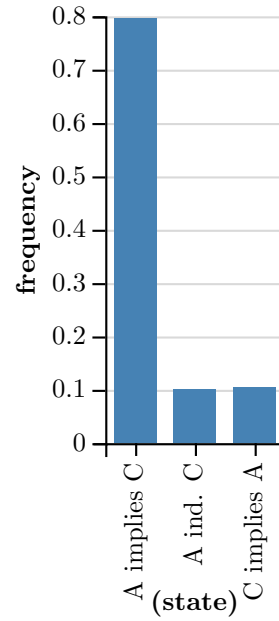
(a) Heard utterance: *A*.



(b) Heard utterance: *C*.



(c) Heard utterance: *If A, C*.



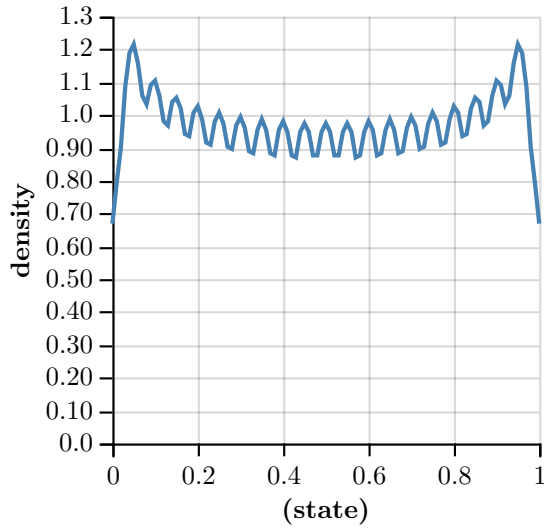
(d) Heard utterance: *If C, A*.

Figure 18: Results for the Pragmatic Listener depending on the utterances he has heard given that the Literal Listener is represented as distribution over Bayes Networks.

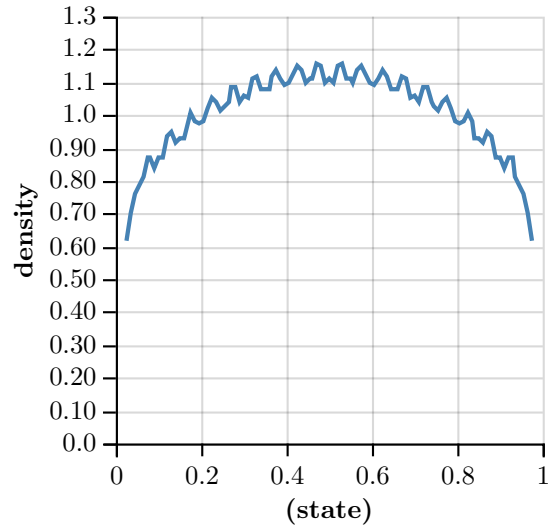
# Appendix

## Literal Listener

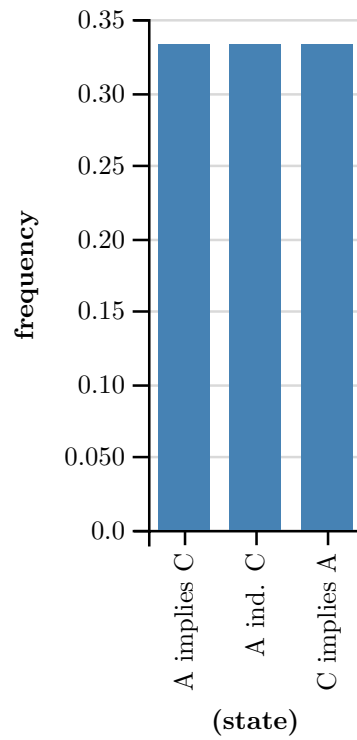
### Prior distributions



(a) Literal Listener's prior for  $P(A|C)$ .



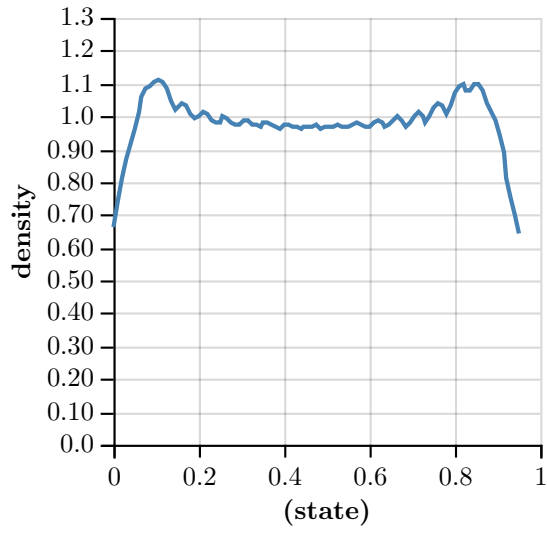
(b) Literal Listener's prior for  $P(A)$ .



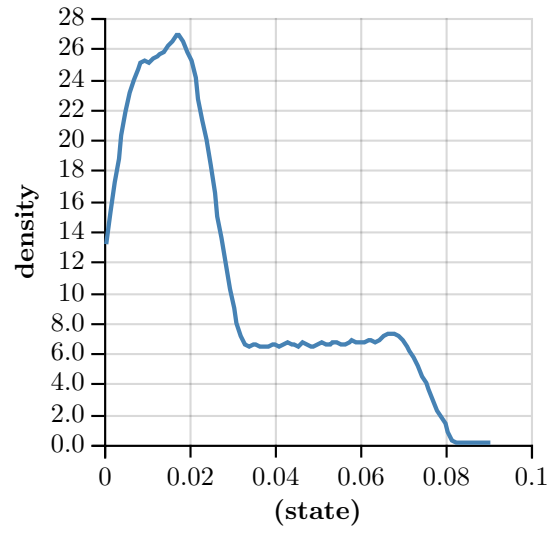
(c) Literal Listener's prior for the Bayes Networks.

Figure 19: Some Prior Distributions of the Literal Listener.

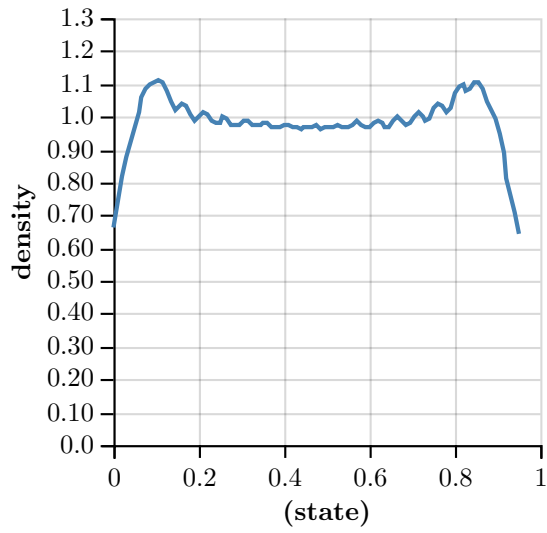
### Joint Probabilities



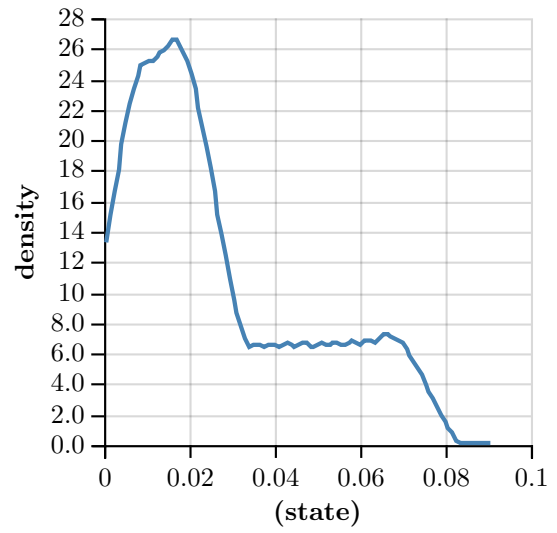
(a)  $A, C$



(b)  $\neg A, C$

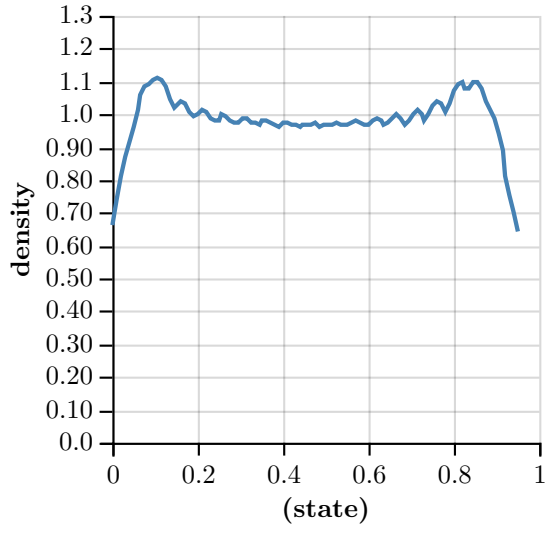


(c)  $A, \neg C$

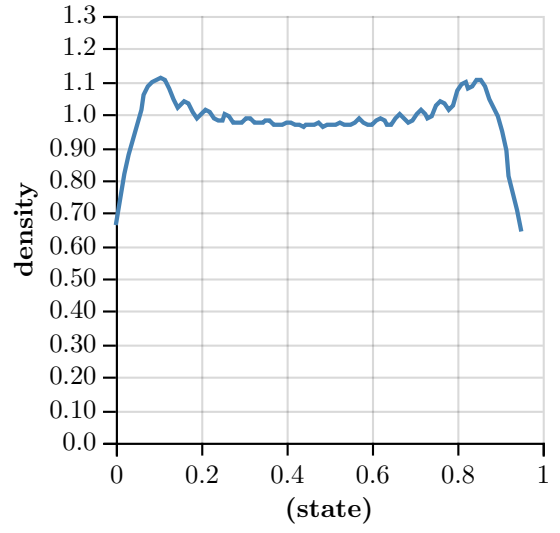


(d)  $\neg A, \neg C$

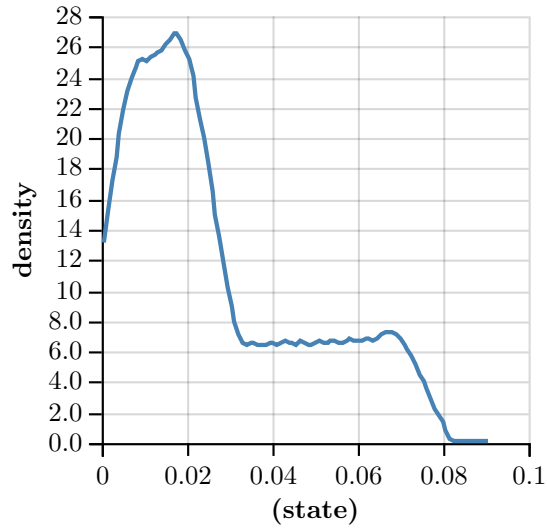
Figure 20: Joint Probability distributions for the Literal Listener when he hears the utterance  $A$ .



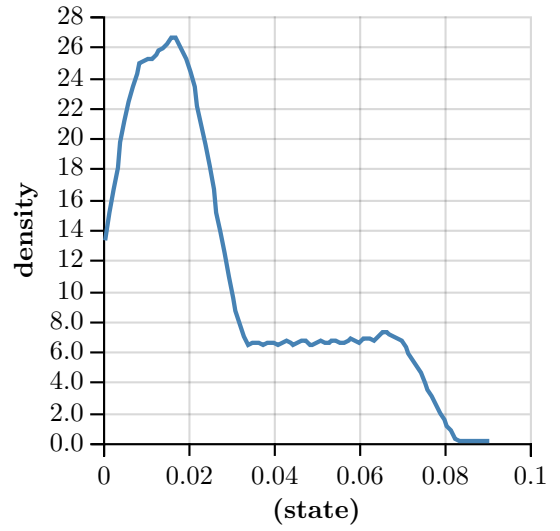
(a)  $A, C$



(b)  $\neg A, C$

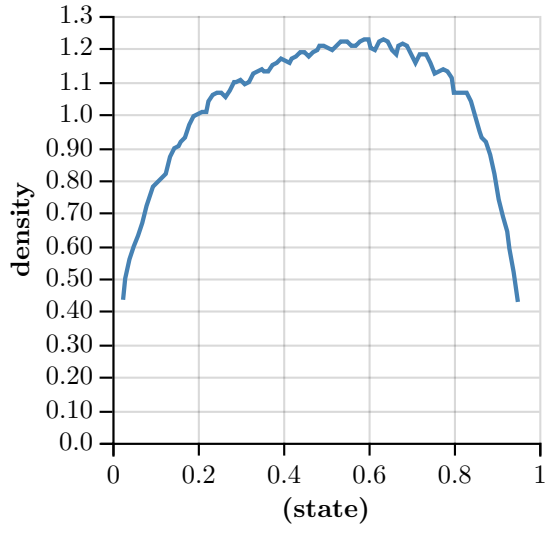


(c)  $A, \neg C$

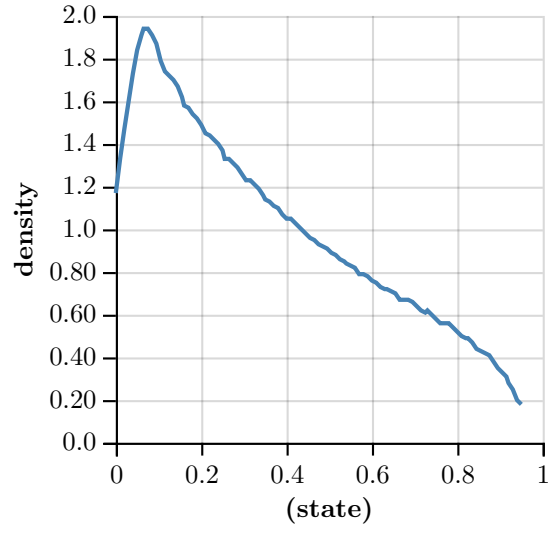


(d)  $\neg A, \neg C$

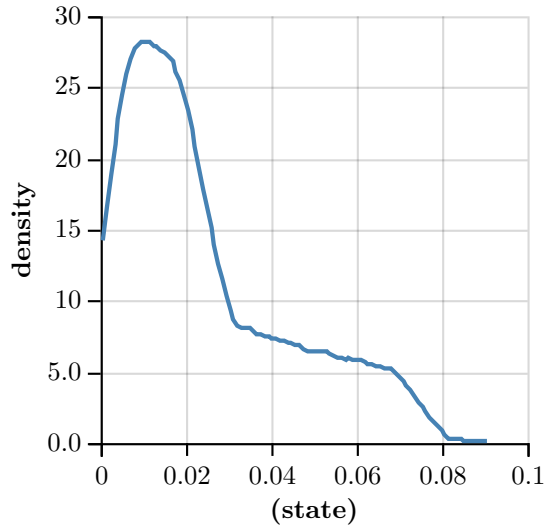
Figure 21: Joint Probability distributions for the Literal Listener when he hears the utterance  $C$ .



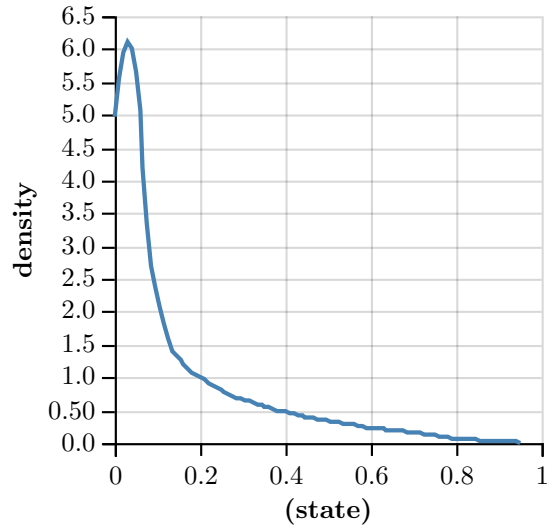
(a)  $A, C$



(b)  $\neg A, C$



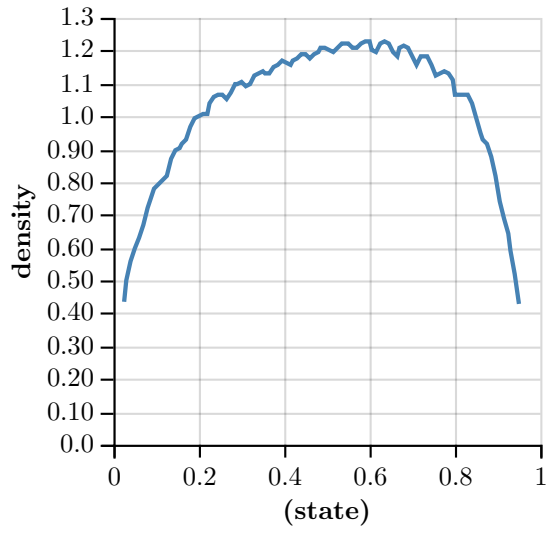
(c)  $A, \neg C$



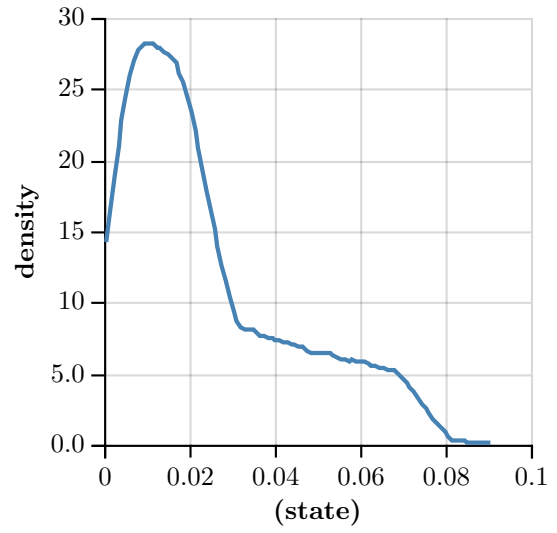
(d)  $\neg A, \neg C$

Figure 22: Joint Probability distributions for the Literal Listener when he hears the utterance *If A, then C*.

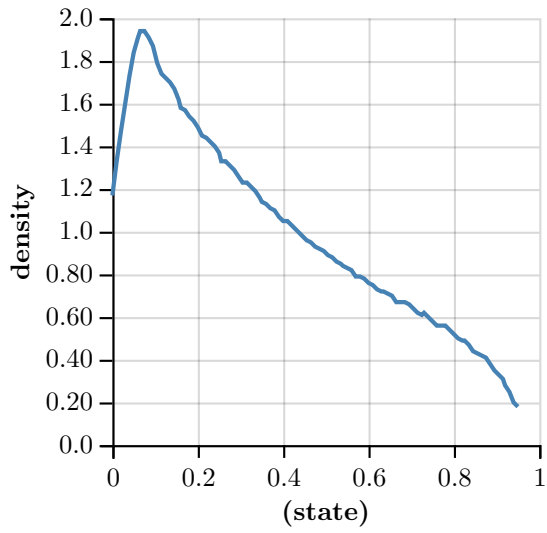




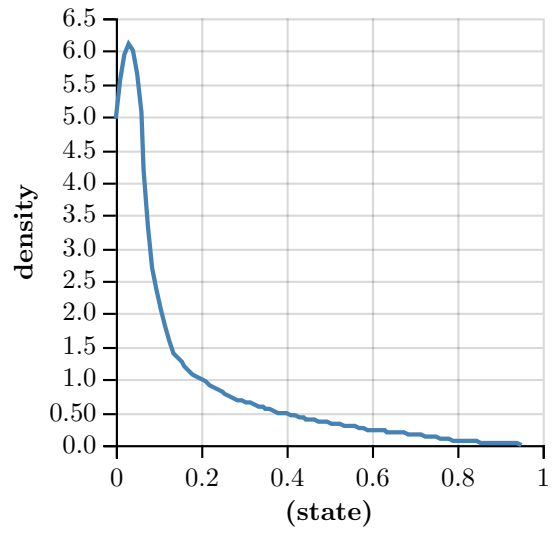
(a)  $A, C$



(b)  $\neg A, C$



(c)  $A, \neg C$



(d)  $\neg A, \neg C$

Figure 23: Joint Probability distributions for the Literal Listener when he hears the utterance *If C, then A*.