Using hidden information is an important problem in agent design. In many environments agents have access to information which is not publically available. This information can be useful as agents decide how best to take advantage of the environment. A complication arises in multiplayer games, when acting on information may reveal that information to the opponent, which can be disadvantageous in a competitive environment. In this paper we will describe the game of Hidden Matching Pennies, in which one player has hidden information which they can use to their profit. We hope that strategies appropriate to this game can be extended to other environments with hidden information.

Hidden Matching Pennies is a game played in stages between two players. In each stage, a game of Matching Pennies is played, with the winner receiving a payoff of 1. In addition, nature chooses a preferred move according to a known probability distribution. One player is informed what the preferred move is. If that player plays the preferred move, they receive an additional payoff of .5 (regardless of whether they win or lose). All payoffs are zero sum, so whenever one player receives a payoff, their opponent receives an equal penalty - this is done to make analysis of the resulting game simpler. By convention, the informed player is always the player trying to match the opponent and the uninformed player is always trying to differ from the opponents move. This choice doesn’t have an effect on the analysis of the game.

Let’s consider what effect this modification has on the game of Matching Pennies. First, let’s look at behavior over a single stage. Here are tables showing the base game, Matching Pennies, the game of Hidden Matching Pennies where Nature chooses Heads as a favored move half the time, and a more generic version of Hidden Matching Pennies where Nature chooses Heads with probability $p$.

<table>
<thead>
<tr>
<th>Opponent</th>
<th>Agent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Heads</td>
</tr>
<tr>
<td>Heads</td>
<td>1</td>
</tr>
<tr>
<td>Tails</td>
<td>-1</td>
</tr>
</tbody>
</table>

Table 1: The Matching Pennies game matrix

<table>
<thead>
<tr>
<th>Uninformed Agent</th>
<th>Agent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Heads</td>
</tr>
<tr>
<td>Heads</td>
<td>1.25</td>
</tr>
<tr>
<td>Tails</td>
<td>-.75</td>
</tr>
</tbody>
</table>

Table 2: The Hidden Matching Pennies game matrix. Nature chooses Heads as a preferred move with 50% probability. Payoffs shown are the expected payoffs, and players are assumed to be risk-neutral.

So let’s begin by considering a p-value of 0 or 1. This is equivalent to a game with no hidden information, as both players will be aware of the favored move. In that situation the Nash equilibrium is for the informed player to play a distribution of (.5,.5) and the uninformed player to play the unfavored move with probability 5/8 (this gives the informed player a choice between playing the favored move and getting the extra payoff (but being more likely to lose) or playing the
Table 3: The Hidden Matching Pennies game matrix. Nature chooses Heads as a preferred move with \( p \) probability. Payoffs shown are the expected payoffs, and players are assumed to be risk-neutral. Only the informed players payoff is shown - the other players' payoff will be the negative unfavored move and being more likely to win). This gives the informed player an expected payoff of .25. This can be viewed as the baseline payoff available to the informed player when there is no hidden information to take advantage of.

At the other extreme is a \( p \)-value of .5. In this situation the uninformed player has the maximum uncertainty about the favored move. The Nash equilibrium is for the informed player to play the favored move, and the uninformed player to play Heads with probability between 3/8 and 5/8. In this situation the informed player has an expected payoff of .5. After this game, the uninformed player will be certain about the favored move.

For intermediate values one move is more likely than the other to be favored. In this situation, the uninformed player will play the move which is more likely to be favored with a probability of 3/8, and the other move with a probability of 5/8. If the favored move is the unlikely option, then the informed player will play the favored move. If the favored move is the likely option, then the informed player will play that move with probability \( \frac{1}{2p} \) where \( p \) is the probability of the more likely option.

Now we have some constraints on the achievable payoffs - the informed player can, at best, achieve an expected payoff of .5 per game (if they can make the opponent forget everything they've seen), and at worst should get an average payoff of .25. Expected payoff for the informed player is linearly related to \( p \).

Dealing with hidden information is fairly straightforward in a single iteration of the game, because there is no need to worry about the value of concealing information for use in the future. With only a single iteration, it doesn't matter if your opponent figures out what the favored move is, and you have no opportunity to deceive them to take advantage of them in the future. So the question is, what happens if you do have a chance to deceive them - what happens over more than one iteration?

The game is simple enough to solve completely over a sufficiently short number of iterations. (In this case, sufficiently short is 5 or fewer repeated games).

0.1 2 Iterations

In 2 iterations, the Nash equilibrium is for the informed player to play randomly (.5,.5) on the first move, and then play the favored move on the second move. The uninformed player will play randomly (.5,.5) on the first move, and on the second move, they will play whatever the first player did with 3/8 probability. In this situation, the uninformed player will not learn the favored move until after the second stage, because the first move of the informed player reveals no information. Since the informed players play does not depend at all on the play of the uninformed player, we
can characterize their strategy as a distribution over sequences of moves made - in this case, (for Heads as a favored move) it’s an even distribution between Heads-Heads and Tails-Heads.

0.2 3 Iterations

In a 3 stage game, the informed player switches between 2 strategies. In the first strategy they choose randomly (.5,.5) between Heads-Tails and Tails-Heads for the first two moves, and play the favored move for the last move. In the second strategy, they play the favored move twice in a row, and then play randomly for the last move. The uninformed player plays randomly the first two moves, and the last move, they play what the first player played twice with probability 3/8. If the first player played different moves on the first 2 moves, they play an arbitrary strategy with probability 3/8. Again, this strategy results in a uniform distribution over play sequences for the informed player. If the favored move is Heads, the possible sequences are Heads-Heads-Heads, Heads-Tails-Heads, Tails-Heads-Heads, and Tails-Heads-Heads.

0.3 More Iterations

<table>
<thead>
<tr>
<th># of Stages</th>
<th>Heads Favor</th>
<th>Expected Final Payoff</th>
<th>Average Per Stage Payoff</th>
<th>Information Value</th>
<th>Per Stage Information Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50%</td>
<td>.5</td>
<td>.5</td>
<td>.25</td>
<td>.25</td>
</tr>
<tr>
<td>2</td>
<td>50%</td>
<td>.75</td>
<td>.375</td>
<td>.25</td>
<td>.125</td>
</tr>
<tr>
<td>3</td>
<td>50%</td>
<td>1.125</td>
<td>.375</td>
<td>.375</td>
<td>.125</td>
</tr>
<tr>
<td>4</td>
<td>50%</td>
<td>1.375</td>
<td>.34375</td>
<td>.375</td>
<td>.09375</td>
</tr>
<tr>
<td>5</td>
<td>50%</td>
<td>1.71875</td>
<td>.34375</td>
<td>.46875</td>
<td>.09375</td>
</tr>
</tbody>
</table>

Table 4: Effect of number of stages on payoffs and the value of hidden information

Outline: Solved completely 1/2/3/4/5 turns Extension to higher numbers Graphs Information value by # turns remaining Information value by # turns remaining and entropy in public info

1 Extension to more moves

There is a clear pattern to Nash equilibrium play for odd numbers of stages, namely, to play an even probability distribution over all sequences of moves in which the favored move occurs more than half the time. It is a bit more complicated for even numbers of stages - in that situation you still play an even distribution, but you discard half of the sequences where you play both moves equally. We have verified this for 5 or fewer stages, now we will extend it to more than 5 stages.

First, let’s note that this fulfills the requirements for a Nash equilibrium. In hidden matching pennies, any equilibrium strategy must involve randomization between heads and tails (for both players) because the payoff for winning the game of matching pennies dominates the payoff for playing the preferred move. Let’s look at things from the perspective of the uninformed player - they would prefer to play a different move from the informed player, and they would prefer that the informed player play the unfavored move. In order for them to play a Nash equilibrium,
they must value each of their moves equally. Since their choice of move cannot affect whether
the informed player plays the favored move, their only interest is not to play the same move as
the informed player. By using Nature's move as a randomization device, the informed player is
effectively playing a uniform distribution over all possible sequences of moves, which will prevent
the uninformed player from having a preference between Heads and Tails. Characterizing the play
of the uninformed player is more complex. If the favored move has been revealed by the prior play
of the informed player, then the uninformed player clearly must play that move with a probability
of 3/8, but it's less clear how the uninformed player should play under uncertainty.

Given that set of strategies of the informed player, we can calculate their expected payoff over
longer sequences of moves. The expected payoff will be

\[
E_{\text{expected}} = E_{\text{of Wins}} - E_{\text{of Losses}} + \frac{1}{2} E_{\text{of Favored}}
\]

Note that in aggregate, the informed player will play heads and tails with equal likelihood. This
means that the expected number of wins should equal the expected number of losses, so we are left with
calculating the expected number of favored move plays. This gives us

\[
E_{\text{of Favored}} = \sum_{h=\frac{N+1}{2}}^{N} h \cdot \left( \begin{array}{c} N \\ h \end{array} \right)
\]

Note that there are \(2^N\) possible sequences, and only half of those will be used, so we have

\[
\sum_{h=\frac{N+1}{2}}^{N} \left( \begin{array}{c} N \\ h \end{array} \right) = 2^{N-1}
\]

Further simplification gets us to

\[
E_{\text{of Favored}} = \frac{N \cdot \frac{2^{N-1} + \left( \frac{N-1}{2} \right)}{2}}{2^{N-1}}
\]

The terms of this equation can be seen as the expected number of favored moves you would
get if you were playing randomly, plus the additional favored moves you can play as a result of
exploiting hidden information.

2 Related Work

Computing Optimal Strategies to Commit to in Extensive Form Games - Letchford and Conitzer
EC 2010

Extensive games and the problem of information H W Kuhn (foundation paper)
<table>
<thead>
<tr>
<th>Number of Stages</th>
<th>Heads Favored</th>
<th>Expected Final Payoff</th>
<th>Average Per Stage Payoff</th>
<th>Information Value</th>
<th>Payoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50%</td>
<td>.5</td>
<td>.5</td>
<td>.25</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>50%</td>
<td>3.42</td>
<td>.311</td>
<td>.676</td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>50%</td>
<td>27.259</td>
<td>.269</td>
<td>2.009</td>
<td></td>
</tr>
<tr>
<td>1001</td>
<td>50%</td>
<td>256.56</td>
<td>.25630</td>
<td>6.312</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Projected effect of number of stages on payoffs and the value of hidden information

The State of Solving Large Incomplete-Information Games, and Application to Poker, Sandholm 2010 - AAAI magazine
Computing approximate Nash equilibria and robust best-responses using sampling Ponsen, de Jong, Lanctot
Successful Performance via Decision Generalization in No Limit Texas Hold’em J Rubin, I Watson
Finding Optimal Abstract Strategies in Extensive-Form Games M Johanson, N Bard, N Burch, M Bowling 2012
Adverse Selection Without Hidden Information, Paul Milgrom
This paper discusses the distinction between hidden information and hidden actions in the context of adverse selection problems. They demonstrate that the fundamental dynamic of an adverse selection problem doesn’t depend on whether it’s a hidden information or a hidden action problem. This tends to confirm my intuition about the similarity of hidden information and simultaneous action, but it isn’t that important as it’s not the primary point of this paper.

Comparing UCT versus CFR in Simultaneous Games, Shafiei, Stutevant, Schaeffer
This paper is useful because it discusses UCT in simultaneous move games. However, their main interest is in pointing out that UCT can be exploited (by CFR, but I don’t think that is essential). The thing to note here is how they handle simultaneous moves. They treat the stage at which simultaneous moves occur as a point at which both players can make a choice, and do so according to the aggregated results observed in prior runs (very similar to fictitious play). Modifying this approach to handle other types of hidden information would result in an approach which disregards the hidden information. Alternatively, you could model it by creating player strategies which map hidden information to actions, but this radically increases the number of actions for each player (doable in test-model cases, for comparison but not fundamentally useful).

Multiple Tree for Partially Observable Monte-Carlo Tree Search
This paper discusses using UCT to play phantom tic-tac-toe (tic tac toe, but you can’t see your opponent’s moves - if you try to make an illegal move, it tells you, and you can try a different spot. They look at using multiple UCT trees which are built simultaneously - each player has it’s own tree, and plays according to it while in a game state in the tree, and otherwise plays randomly. As you do exploration, you build all the trees at once. This is supposed to provide convergence to a NE. They compare performance of this approach with various numbers of iterations to random play, and to a bayesian approach. The bayesian approach does fairly well - especially at exploiting weaker players. I’m not sure how useful this approach is, but I don’t think I like it that much. For one thing, it seems like the trees would end up being built in the same ways.

Monte-Carlo Tree Search in Poker using Expected Reward Distributions
Uses Monte-Carlo tree search, but handles opponent actions as random events, and not as
Mixing Search Strategies for Multi-Player Games, Zuckerman, Felner, Kraus

Looks at balancing between a MaxN and Paranoid approach, depending on the circumstances of the game. This will need a closer look

Combining Online and Offline Knowledge in UCT, Gelly, Silver

Modifies UCT by using pre-calculated strategies for the play-out phase

Multi-player Go, Cazenave

This paper looks at modifying UCT to play 3-player go. It’s methodology is to try different evaluation functions and playout policies, and run them against each other to see which approaches perform better. It is somewhat related to the problem of hidden information, in that in a multiplayer game there is uncertainty about the actions undertaken by the other players, since there are issues with how players will cooperate with each other. They answer this question by building multiple trees with different cooperation patterns, and selecting the final move based on the properties of the different trees. It is not clear how useful this approach will be for me, since the space of the hidden information in their situation (opponent intent, modelled as a binary choice of cooperation partners) is so severely restricted. On the other hand, the beginning of their approach is similar to what I’ve been thinking of - namely, a playout assuming full public information.

### 3 Graphs

<table>
<thead>
<tr>
<th>Player 2</th>
<th>Player 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>2.99 - 1.0150.76</td>
</tr>
<tr>
<td>2000</td>
<td>2.98 - 1.0250.78</td>
</tr>
<tr>
<td>4000</td>
<td>2.98 - 1.0550.72</td>
</tr>
<tr>
<td>8000</td>
<td>2.98 - 1.0650.72</td>
</tr>
<tr>
<td>16000</td>
<td>2.98 - 1.0650.72</td>
</tr>
</tbody>
</table>

Table 6: Mutual Hidden Matching Pennies outcomes for game length 3

<table>
<thead>
<tr>
<th>Player 2</th>
<th>Player 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>3.87 - 1.0350.18</td>
</tr>
<tr>
<td>2000</td>
<td>3.91 - 0.9850.13</td>
</tr>
<tr>
<td>4000</td>
<td>3.91 - 0.6250.53</td>
</tr>
<tr>
<td>8000</td>
<td>3.93 - 1.0450.13</td>
</tr>
<tr>
<td>16000</td>
<td>3.94 - 1.4350.88</td>
</tr>
</tbody>
</table>

Table 7: Mutual Hidden Matching Pennies outcomes for game length 4
4 Mathematical Notes

$h = \text{probability favored move is heads.}$
$t = \text{probability of terminating after a stage}$

$P_H(h, t) = \text{Probability informed player will play Heads given Heads is the favored move}$

$P_T(h, t) = \text{Probability informed player will play Heads given Tails is the favored move}$

$P_U(h, t) = \text{Probability uninformed player will play Heads}$

$V_H(h, t) = \text{Value of future play (assuming there is future play) if favored move is Heads}$

$V_T(h, t) = \text{Value of future play (assuming there is future play) if favored move is Tails}$

$V(h, t) = hV_H(h, t) + (1 - h)V_T(h, t)$

$hP_H(hmt) + (1 - h)P_T(h, t) = .5$

$V_H(0, t) = \frac{1}{7}$

$V_T(0, t) = \frac{3}{7}$

$V_H(1, t) = \frac{1}{7}$

$V_T(1, t) = \frac{3}{7}$

$\frac{2}{5} \leq P_U(h, t) \leq \frac{5}{8}$

$P_H(h, t) = P_T(1 - h, t)$
4.1 Theories

The informed agent will always play the favored move with probability greater than or equal to the unfavored move. (This does not mean no bluffing, but it does imply something about the frequency of bluffing).

If the unlikely move is favored, the informed agent will play the favored move with probability 1 if there is only one move left.

The informed agent will play the favored move with greater probability if the favored move is the unlikely move.

The uninformed player will always play the unlikely move with probability greater than or equal to the likely move.

In a fixed-length sequence, the informed agent will (in combination with Nature) always play a uniform distribution over all possible sequences.

\[
V_H(h, t) = P_H(h, t)(2P_U(h, t) - 1) + (1 - t)V_H(h^{\frac{hP_H(h, t)}{hP_H(h, t)+(1-h)P_T(h, t)}}, t) + (1 - P_H(h, t))(1 - 2P_U(h, t) + (1 - t)V_H(h^{\frac{h(1-P_H(h, t))}{h(1-P_H(h, t))+(1-h)(1-P_T(h, t))}}, t))
\]
In a random-length sequence of play, the informed agent will play a uniform distribution over all possible sequences.
Table 8: Mutual Hidden Matching Pennies outcomes for game length 5
Figure 5:

Table 9: Mutual Hidden Matching Pennies outcomes for game length 10
Figure 6:

Table 10: Mutual Hidden Matching Pennies outcomes for game length 20

<table>
<thead>
<tr>
<th>Player 2</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
<th>8000</th>
<th>16000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fail</td>
<td>Fail</td>
<td>Fail</td>
<td>Fail</td>
<td>Fail</td>
<td>Fail</td>
</tr>
<tr>
<td>Fail</td>
<td>Fail</td>
<td>Fail</td>
<td>Fail</td>
<td>Fail</td>
<td>Fail</td>
</tr>
<tr>
<td>Fail</td>
<td>Fail</td>
<td>Fail</td>
<td>Fail</td>
<td>Fail</td>
<td>Fail</td>
</tr>
<tr>
<td>Fail</td>
<td>Fail</td>
<td>Fail</td>
<td>Fail</td>
<td>Fail</td>
<td>Fail</td>
</tr>
<tr>
<td>Fail</td>
<td>Fail</td>
<td>Fail</td>
<td>Fail</td>
<td>Fail</td>
<td>Fail</td>
</tr>
</tbody>
</table>
Figure 7: