If You Can't Be With the One You Love, Love the One You’re With: How Individual Habituation of Agent Interactions Improves Global Utility
Davies et al 2010

Presented by Peter Helfer
EGT Reading group 2012-11-23
Catchy title

• Lesson learned:
  – A catchy title can increase your readership.

• Or:
  – To get the attention of aging baby boomers, use a 1970s rock and roll song title.
Myopic agents optimizing Global Utility

• Engineered systems often use components with simple local processing rules to optimize global performance.
  – Example: telecom networks

• Some biological systems appear to do the same
  – Example: some symbiotic species (?)

• How is this possible?
More specifically

• Given a network of “selfish” agents that strive to maximize their own utility:

• If it is observed that collectively these agents create adaptations that cannot be explained as the result of such selfish behavior,

• Then it must be the case that sometimes agents make decision that are detrimental to their self-interest.
Question

• How can it make sense to talk about selfish agents that make decisions that do not further their self-interest?
Reality Distortion

• Suppose that it is possible to introduce a small distortion into the agents’ perception of the utility of their choices,
• such that they unwittingly further the common good,
• in their misguided attempts to be selfish.
Habituation

• The authors claim that “habituation” is such a distortion:
• A tendency to keep making the same choices,
• i.e. an increased preference for making the same choice again, regardless of the outcome,
• will result in an optimization of global utility.
• Yes, this is quite surprising.
Example: Drinking Scientists

• Consider a network of scientists, with a symmetric “compatibility” relationship:
  – the higher the compatibility between two individuals, the more productive it is for them to collaborate.

• Further assume that all collaborations take place in one of two bars, A and B
• Each night each scientist (in random order) decides whether to go to the same bar as the previous night, or switch to the other bar, based only on which colleagues are currently at each of the two bars.

• Each scientist collaborates with all the other scientists in the bar and reaps productivity accordingly.

• Since compatibilities are symmetric, an equilibrium will soon be reached where no-one wants to switch pubs.

• So the equilibrium is a local maximum of global productivity.

• But in general, this equilibrium will not maximize total productivity.
Habituation

- Introduce a preference for drinking with the same colleagues you drank with last night (not necessarily a preference for staying at the same bar.)

- The (claimed) result:
  - Over many iterations, the scientists will gradually settle on a configuration (distribution over A and B) that maximizes (or nearly maximizes) total productivity.
Questions

1. Why does habituation promote global utility? How does this work?

2. Why would habituation – a preference for status quo over maximization of utility – ever evolve (i.e. be selected for)?
The Model

- N=100 agents playing two-player games on a fully connected network.
- At each iteration, each player plays against all the others, adopting either strategy A or B towards all of them.
- The payoff matrix:

<table>
<thead>
<tr>
<th></th>
<th>Player 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>$\alpha,\alpha$</td>
</tr>
<tr>
<td>B</td>
<td>$\beta,\beta$</td>
</tr>
</tbody>
</table>

Coord: $\alpha = 1, \beta = 0$
Anti-coord: $\alpha = 0, \beta = 1$

- Each pair of agents (i,j) play either the coordination or the anti-coordination game, with equal probability.
Note

- This model is NOT exactly like the drinking scientists example:
  - Scientists’ payoff matrices:
    - Player payoff matrices:
      \[
      \begin{array}{c|cc}
      \text{Player } i & \text{A} & \text{B} \\
      \hline
      \text{A} & \alpha_{ij} & 0 \\
      \text{B} & 0 & \alpha_{ij} \\
      \end{array}
      \]
      \[
      \alpha_{ij} \geq 0.0
      \]
  - Model’s payoff matrices:
    - Player payoff matrices:
      \[
      \begin{array}{c|cc}
      \text{Player } i & \text{A} & \text{B} \\
      \hline
      \text{A} & 1 & 0 \\
      \text{B} & 0 & 1 \\
      \end{array}
      \]
      or
      \[
      \begin{array}{c|cc}
      \text{Player } i & \text{A} & \text{B} \\
      \hline
      \text{A} & 0 & 1 \\
      \text{B} & 1 & 0 \\
      \end{array}
      \]
And by the way…

• Wouldn’t the scientists maximize their productivity by simply all going to same bar?

• Not very interesting…

• Or am I missing something?
but let’s return to the model…
Utility

At each iteration, agent i’s utility is:

\[ u_i(t) = \sum_{j}^{N} U_{ij}(S_i(t), S_j(t)) \]

where:

- \( U_{ij}(S_i(t), S_j(t)) \) is the payoff received from playing against agent j at time t.
- \( S_n(t) \) is the strategy played by agent n at time t.

At each iteration, a purely selfish agent would decide to play \( s_i(t) = A \) or \( B \), depending on which strategy maximizes \( u_i(t) \).
Global Utility

Global utility is simply the sum of individual utilities:

\[ G(t) = \sum_{i}^{N} \sum_{j}^{N} U_{ij}(S_i(t), S_j(t)) \]
Habituation

• In addition to the “true” utility matrices $U_{ij}$, each agent possesses a set of “preference” matrices $P_{ij}$.

• Instead of using the true utility, the habitual agent chooses its strategy based on the perceived utility:

$$p_{i(t)} = \sum_{j}^{N} \left[ U_{ij}(s_{i(t)},s_{j(t)}) + P_{ij}(s_{i(t)},s_{j(t)}) \right]$$

• All the preference matrices are initially all-zeros, but very gradually change to reinforce a preference for status quo.
Preference adjustment

- At each time step, an agent’s set of preference matrices is updated with probability 0.0001. The update of each $P_{ij}$ is performed thus:

- Update matrix $A =$

$$
\begin{array}{|c|c|c|}
\hline
\text{Player } i & \text{A} & \text{B} \\
\hline
\text{A} & r & -r \\
\text{B} & -r & r \\
\hline
\end{array}
$$

- Compute $P'_{ij} =$ either $P_{ij}(t) + A$ or $P_{ij}(t) - A$ (with equal probability).
- If $P'_{ij}$ increases $p_i(t)$, then $P_{ij}(t+1) = P'_{ij}$
In English:

- Once every 10,000 time steps an agent updates roughly half its preference matrices to slightly increase its preference for the strategy it is currently playing against those agents.

- (A simulation consists of 1,000,000 time steps.)
The Experiment

• One “relaxation” consists of:
  – Randomize all the strategies $S_i$
  – Run 1000 time steps, in which all the agents (in random order) are allowed to choose strategies.
  – Record the final global utility $G$.

• One run:
  – Run 1000 relaxations without habituation, followed by 1000 relaxations with habituation.
Results
A few sample relaxations without habituation
A few sample relaxations with habituation
Relaxation end points with/without habituation
Global utility with/without habituation
So…

- This demonstrates the authors’ claim: Habituation does promote global utility.

- Or, put another way:
  - “Over time agents come to choose behaviors that decrease their individual utility in certain circumstances, but that allow the system to ultimately reach states of global utility higher than would have been otherwise possible.”
Why does it work?
Analysis

An alternative formulation is to say that:

- all the games are cooperation games
- the edges have weights \( w_{ij} = +1 \) or \(-1\)
- the payoffs are \( w_{ij} * U_{ij} \)
- Now, add in the preference matrix:

\[
\text{payoff} = (w_{ij} + k_{ij}r) * U_{ij}
\]

where

- \( w \) is the original weight
- \( r \) is the preference delta
- \( k_{ij} \) is the number of times i & j have coordinated to date.
This perspective makes it clear that the simulation is equivalent to Hebbian learning on a discrete Hopfield net (recurrent neural network).

<table>
<thead>
<tr>
<th>Strategy</th>
<th>↔</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{ij} = \text{True utility } + \text{preference}$</td>
<td>↔</td>
<td>Connection weight</td>
</tr>
<tr>
<td>Preference delta ($r$)</td>
<td>↔</td>
<td>Learning rate</td>
</tr>
<tr>
<td>Global utility</td>
<td>↔</td>
<td>-Energy</td>
</tr>
</tbody>
</table>
Energy Landscape
But what is the network learning?

There are no externally imposed “training patterns”…

Answer: it is learning those patterns in which it spends most of its time under its intrinsic dynamics,

i.e. the local maxima of the global utility function.

Well, this was shown empirically in the simulation, but not – as far as I could tell – justified theoretically.
In other words,

The system is learning its own attractors.

But what does this mean?

It means that the basins of frequently visited attractors grow wider, and eventually one attractor will out-compete the others and its basin will grow to encompass the entire state space.

But why would the winning attractor be the one corresponding to the highest global utility?
Well, that’s just the way it is…

Or, more precisely, it has been proven that in a system of this kind – a system composed of a superposition of many pairwise interactions – the area of an attractor is positively correlated to its depth.

In other words, the global maximum will often have the largest basin, and so will win most often, and thus be learned most strongly.
And furthermore…

Hopfield nets tend to develop “spurious attractors”, corresponding to superpositions of the trained patterns.

In neural network such attractors are a nuisance, but here they actually help the development of the global-optimum attractor, which necessarily shares many behavior pairs with all the local optima…

Again, it is not clear why the energy minimum in the Hopfield net corresponds to the *true* utility of the simulation.
So what about that second question?

Why would the trait of habituation be selected for?

The paper does not address this question, except to say that “habituation is ‘intuitively’ natural for any agent that favors exploitation over exploration – as any non-teleological agent must”.

Conclusion

• It’s interesting to see the mathematical formalism from one field (here, neural networks) be applied to another (game theory).

• Is there any deep meaning in this?

• Probably not, but there is potential for new discoveries, as when the math of statistical mechanics was borrowed into neuroscience (Hopfield), or that of quantum mechanics into decision theory (Busemeyer).
The End