

# Lernen in HUME<sub>1.0</sub>

April 11, 2008

## 1 Randbedingungen

Im folgenden versuche ich, einen Lernmechanismus für HUME<sub>1.0</sub> zu geben. Der Mechanismus ist zunächst einmal lediglich auf das PM-Szenario zugeschnitten. Das Lernen betrifft in diesem Szenario vier Wahrscheinlichkeiten, nämlich

1. als P-Agent auf dem Markt nach einem S-Agenten zu suchen (und mit der Gegenwahrscheinlichkeit innerhalb der lokalen Nachbarschaft zu agieren);
2. als S-Agent auf dem Markt seine Problemlösungskompetenz anzubieten (und mit der Gegenwahrscheinlichkeit innerhalb der lokalen Nachbarschaft zu verbleiben);
3. als S-Agent bei lokaler Interaktion *rewarder* zu sein (und mit der Gegenwahrscheinlichkeit *exploiter* zu sein);
4. als S-Agent auf dem Markt *rewarder* zu sein (und mit der Gegenwahrscheinlichkeit *exploiter* zu sein).

Formal ist damit ein Agent  $i$  zum Zeitpunkt  $t$  durch den folgenden Entscheidungsvektor charakterisiert:

$$M_i^{PM}(t) = \langle p_i^{P \rightarrow \text{market}}(t), p_i^{S \rightarrow \text{market}}(t), p_i^{\text{reward\_local}}(t), p_i^{\text{reward\_market}}(t) \rangle$$

## 2 Lernprobleme

Die Veränderungen der vier Komponenten erfolgen durch Lernen. Bei diesem Lernen gibt es Alternativen. Ein naheliegender Ansatz ist der Payoff-Vergleich mit anderen, und zwar – jedenfalls zunächst – unter der idealisierenden Voraussetzung einer *Sichtbarkeit* sowohl der Payoffs wie der Wahrscheinlichkeiten im Entscheidungsvektor.

Die Sichtbarkeitsannahme ist *nicht unproblematisch*: Zum einen könnten Agenten ein Interesse daran haben, ihre niedrige Wahrscheinlichkeiten in den Komponenten  $p_i^{\text{reward\_local}}(t)$  bzw.  $p_i^{\text{reward\_market}}(t)$  nicht offenbar werden zu lassen und dabei in einem bestimmten Umfange zu Täuschung in der

Lage sein. Zum anderen läßt sich bezüglich  $p_i^{\text{reward\_market}}$  nur schwer argumentieren, die Sichtbarkeitsannahme sei eine Art *short cut*, der die explizite Modellierung eines Reputationsmechanismus erspart.

Auch unter der idealisierenden Sichtbarkeitsannahme stellen sich dann zahlreiche weitere Fragen:

1. Gibt es nur eine globale oder aber eine differenzierte, nach Ort und Art der Interaktion unterscheidende Buchführung (Markt, Nachbarschaft, rewarder, exploiter)?
2. Werden Payoffs einfach aufsummiert oder werden vergangene Auszahlungen abdiskontiert?
3. Wie wird berücksichtigt, daß S-Agenten als exploiter klassifiziert werden und gegebenenfalls beim Matching nicht gepaart werden?
4. Wo wird gelernt? In der Nachbarschaft, auf dem Markt oder an beiden Orten?
5. Wer lernt von wem? Orientiert man sich an jemandem mit den *besten* Auszahlungen (*role model* bei Macy/Sato)? Werden für das Lernen *zufällig* Agenten ausgewählt bzw. gepaart, die dann ihre Propensities im Lichte ihrer Payoffs modifizieren (stilisierte Treffen am Lagerfeuer bzw. – später – in einer Bar)? Oder wertet man – ganz ohne Vergleich mit anderen – nur die *eigenen* Payofferfahrungen unter Heranziehen von Anspruchsniveaus aus?
6. Was wird gelernt? Nur hinsichtlich einer Komponente des Entscheidungsvektors? Oder gleich hinsichtlich mehrerer oder sogar aller Komponenten?
7. Wie häufig wird gelernt? In jeder Periode oder jeweils nur mit einer bestimmten Wahrscheinlichkeit?

Wie es scheint, lassen sich zahlreiche alternative (und *prima facie* gleichermaßen plausible) Lernmechanismen denken. Dies wirft Robustheitsprobleme auf, die später zu klären sein werden.

### 3 Ein erster Mechanismus

Die Annahme, daß die Wahrscheinlichkeiten des Entscheidungsvektors und die Payoffs *sichtbar* sind, legt nahe anzunehmen, daß Lernen *nur* innerhalb der lokalen Nachbarschaft erfolgt. Dort ist die Sichtbarkeit von  $p_i^{\text{reward\_local}}$

am ehesten plausibel. Zugleich besteht lokal kein Anreiz, über  $p_i^{\text{reward\_market}}$  zu täuschen. Auch bezüglich der Payoffs läßt sich bei lokalem Lernen am ehesten die Sichtbarkeitsannahme verteidigen.

Was die Aggregation von Payoffs betrifft, so liegt nahe, die bis Periode  $t$  kumulierte Auszahlung  $\Pi_i(t)$  des Agenten  $i$  als Summe der Auszahlung in Periode  $t$  plus der mit  $\gamma$  gewichteten vergangenen Auszahlungen gemäß

$$\Pi_i(t) = \pi_i(t) + \gamma \cdot \Pi_i(t-1)$$

mit  $0 \leq \gamma \leq 1$  zu bestimmen. Für  $\gamma = 1$  zählt die gesamte Vergangenheit mit gleichem Gewicht wie die Gegenwart; für  $\gamma = 0$  zählt nur die Gegenwart. Im Modell sollte  $\gamma$  ein frei wählbarer Parameter sein.

Ein erster einfacher Lernmechanismus sollte sich an das Rollenmodell-Lernen, das auch in Macy/Sato genutzt wird, anlehnen:

1. Aus der Nachbarschaft wird in jeder Periode  $t > 1$  ein Agent  $j$  mit der jeweils größten kumulierten Auszahlung  $\max [\Pi_i(t)]$  (Rollenmodell) ausgewählt. (In Macy/Sato wird das Rollenmodell aus der gesamten Gruppe der *local neighbors* – das sind diejenigen innerhalb der eigenen Gruppe, die schon mindestens eine Periode zu ihr gehören – als Rollenmodell ausgewählt.)
2. Ein lernender Agent  $i$  übernimmt aus dem Entscheidungsvektor des Rollenmodells  $j$  mit einer exogen vorgegebenen Wahrscheinlichkeit  $\beta$  jeweils (einzeln und separat!) die Wahrscheinlichkeit der vier  $j$  charakterisierenden Komponenten. (In Macy/Sato ist  $\beta = 0,5$ .)
3. In jeder Periode  $t$  gibt es ‘*Mutationen*’: Mit einer Wahrscheinlichkeit  $\mu$  wird jede der vier Komponenten des Entscheidungsvektors aller Agenten um einen Betrag  $\delta$  nach oben oder unten verschoben. Die Verschiebungsrichtung sei gleichwahrscheinlich. Wird durch Verschiebung der Wertebereich  $[0,1]$  verlassen, wird eine Korrektur vorgenommen und der mutierte Wert auf 0 bzw. 1 gesetzt. Die Mutationen können als Experimentieren oder auch als Fehlwahrnehmungen bzw. Fehlfortschreibungen eigener und anderer Wahrscheinlichkeiten gedeutet werden. (In Macy/Sato gibt es dieses Element allerdings *nicht*.)

So einfach dieser Lernmechanismus ist, es sind immerhin bereits vier Parameter, die ihn steuern:  $\gamma$ ,  $\beta$ ,  $\mu$  und  $\delta$ . Ein weiterer Parameter käme hinzu, wenn man annimmt, daß das Rollenmodell nicht aus der gesamten Nachbarschaft bestimmt wird, sondern sich in jeder Periode in jeder Nachbarschaft kleinere Gruppen bilden (überlappend oder partitioniert) in denen dann jeweils das Rollenmodell kleingruppenspezifisch bestimmt wird.