▸ FHI TECHNICAL REPORT ◂

# Nash equilibrium of identical agents facing the Unilateralist's Curse


Stuart Armstrong



Technical Report #2012-3



*Cite as:*

Armstrong, S. (2012): "Nash equilibrium of identical agents facing the Unilateralist's Curse", *Technical Report* #2012-3, Future of Humanity Institute, Oxford University: pp. 1-5.

# Nash equilibrium of identical agents facing the Unilateralist's Curse

Stuart Armstrong,
Future of Humanity Institute,
Department of Philosophy,
Oxford University

2012

**Abstract**

This paper is an addendum to the 'Unilateralist's Curse' paper of Nick Bostrom, Thomas Douglas and Anders Sandberg [BDS12]. It demonstrates that if there are identical agents facing a situation where any one of them can implement a policy unilaterally, then the best strategies they can implement are also Nash equilibriums. It also notes that if this Nash equilibrium involves probabilistic reactions to observations, then it is a weak Nash equilibrium and a single agent is free to change all their non-trivial probabilistic decisions, without changing the expected utility of the outcome.

The Unilateralist's Curse paper analyses how to make decisions when there is a certain policy under consideration, and many different agents who could each unilaterally implement that policy. If each agent simply followed their own estimate's of the value of that policy, we would be in a situation similar to the winner's curse in auctions: the policy would get implemented if the *most optimistic* agent thought it was a good idea. Thus in these situations, agents must take care to construct a decision process that counteracts this effect and makes the agents less likely to go ahead on personal, marginally optimistic, information. The problem is isomorphic, in reverse, to policies that require unanimity: there the policy's implementation is dictated by the opinion of the most pessimistic agent.

This paper looks at a specific simplified version of this problem. It assumes that all the agents have identical preferences (they judge each outcome as equally good or equally bad), that they are equally likely to see any given piece of evidence about the value of the policy, and that they can't communicate. They will attempt to construct the best (probabilistic) strategy they can, given these constraints. Because they are identical, they will all construct the same probabilistic strategy. This paper demonstrates that if this is indeed the best strategy (or even a local maxima), then it is a Nash equilibrium: it cannot be improved by unilateral changes by a single agent.

If the strategy is probabilistic (given certain observations, the agent is neither entirely certain to implement the policy, nor entirely certain to refrain), then it is a weak Nash equilibrium – a single agent can change their strategy without making the situation worse. Indeed, a single agent can change all the non-trivial probabilities in their strategy (those neither zero nor one), without changing the expected utility at all.

Before proving those results, we will first set up the problem in the more general situation.

# 1 Identical preferences, no communication

Let there be $n$ agents, and let $\mathbb{O}$ be the (finite) set of possible observations that each agent could make. Each agent will react to the observations in one of two ways: by doing nothing, or by going ahead and implementing the policy unilaterally. They may also choose to implement the policy with a certain probability.

Hence a probabilistic strategy for agent $i$ is a function $f_i : \mathbb{O} \to [0, 1]$, a function from the set of possible observations to a probability value. Then $f_i(o) = p$ means that if agent $i$ observes $o$, she will go do nothing with probability $p$. The set of probabilistic strategies for each agent is a vector $(f_1, f_2, \ldots f_n)$.

Let $\mathbb{W}$ be set of possible states of the world before the agents make their decisions. For $w \in \mathbb{W}$, each agent has a utility $u_0(w)$ (representing their evaluation of no agent going ahead in state $w$) and $u_1(w)$ (representing their evaluation of at least on agent going ahead in state $w$) – by assumption, these utilities are the same for every agent.

The agents will all make their observations. The set of all possible observations is simply a list of $n$ observations, a vector with $n$ elements, each in $\mathbb{O}$ (mathematically, this vector is an element of $\mathbb{O}^n$).

If the agents implement the probabilistic strategies $f = (f_1, f_2, \ldots, f_n)$ and see the observations $O = (o_1, o_2, \ldots, o_n)$, then let $Not(f, O)$ be the probability that no agents do anything: the policy isn't implemented. This is simply the product of all the $f_i(o_i)$ terms, which represent the probability that agent $i$ won't do anything. Thus their expected utility in world $w$ will be:

$$EU(f, w, O) = u_0(w)Not(f, O) + u_1(w)(1 - Not(f, O))$$

In each world-state $w$, there is a certain prior probability for each observation set $O$, $P(O|w)$. Thus in world-state $w$, given strategies $f$, the agents would have an expected utility of

$$EU(f, w) = \sum_{O \in \mathbb{O}^n} P(O|w)EU(f, w, O).$$

Finally, each world-state $w$ has an initial prior probability $P(w)$, ensuring that the expected utility of the strategies $f$ is

$$EU(f) = \sum_{w \in \mathbb{W}} P(w)EU(f, w).$$

In gory detail, this sum is

$$\sum_{w \in \mathbb{W}} P(w) \left( \sum_{O \in \mathbb{O}^n} P(O|w) \Big( u_0(w)Not(f, O) + u_1(w)(1 - Not(f, O)) \Big) \right). \quad (1)$$

It is important to note that though $EU(f)$ is not linear in the components of $f$ (because of the product term $Not(f, O)$), it is linear in the components of any single $f_i$ (since the $Not(f, O)$ is a product of the $f_i(o_i)$'s for all distinct $i$'s).

Each $f_i$ is a vector of probabilities, which are elements of $[0, 1]$. Thus we can differentiate $f_i$ and hence $EU(f)$ in terms of changes in these elements. Even if $f_i(o_i)$ is zero or one (which means that we cannot meaningfully have lower/higher probabilities), we can still differentiate $EU(f)$ as a formal function in terms of changes in the components of $f_i$ – equation (1) doesn't 'know' that the elements of $f$ are supposed to be in $[0, 1]$.

What is a change in $f_i$? Well, if $g_i$ is another possible strategy for agent $i$, then the change in strategy is simply the subtraction $\delta = f_i - g_i$, i.e. for any observation $o$, $\delta(o) = f_i(o) - g_i(o)$. A change in overall strategy $f$ is a vector of such changes in individual $f_i$'s. We'll write $\delta^i$ as the change vector $(0, \ldots, 0, \delta, 0, \ldots, 0)$, with $\delta^i$ being $\delta$ in the $i$th slot and zero everywhere else. Hence $\delta^i$ corresponds to changing the strategy $f_i$ (by adding $\delta$) and leaving the other strategies unchanged.

Because $EU(f)$ is linear in the $f_i$'s, differentiating in the $\delta^i$ direction gives us the exact changes in $EU$, not just a limiting expression. In symbolic form, writing $D(EU)$ as the derivative of $EU$,

$$EU(f + \delta^i)) = EU(f) + (D(EU))_f(\delta^i). \tag{2}$$

## 2 Indistinguishable agents, no communication

We will now consider that the agents are fully identical (in terms of the observations they expect to make, as well as their preferences) and that still don't communicate. Then their combined strategy will be of the form $< g >= (g, g, \ldots, g)$ for some $g$: they will all implement the same strategy.

Then we call $g$ a local maximum symmetric strategy if $EU(< g >)$ is a local maximum in terms of $g$. Note that this does not mean that $< g >$ is a local maximum for $EU$ in the space of all strategies, including the non-symmetric ones. It is possible (indeed plausible) that if we could assign different roles to different agents, we could reach higher expected utility.

Then we have the following encouraging result:

**Theorem 2.1.** *If $g$ is a local maximum symmetric strategy for $EU$, then $< g >$ is a Nash equilibrium for all agents.*

*Proof.* The proof derives from this essential Lemma:

**Lemma 2.2.** *If $\delta$ is a change in strategy, chosen so that $g + \delta$ is also a strategy (i.e. the probabilities in $g + \delta$ remain between 0 and 1), then for any $i$:*

$$(D(EU))_{<g>}(\delta^i) = (D(EU))_{<g>}(< \delta/n >).$$

*Proof of Lemma.*

Since the agents are identical, for all $i$ and $j$, we must have:

$$(D(EU))_{<g>}(\delta^i) = (D(EU))_{<g>}(\delta^j).$$

Since the derivative is linear, we must have

$$(D(EU))_{<g>}(<\delta/n>) = \sum_{j=1}^{n}(D(EU))_{<g>}(\delta^j/n).$$

Replacing the $j$'s with a given $i$:

$$\begin{aligned}(D(EU))_{<g>}(<\delta/n>) &= (D(EU))_{<g>}(\sum_{j=1}^{n}\delta^i/n) \\ &= (D(EU))_{<g>}(\delta^i).\end{aligned}$$

■

The assumption that $g$ is a local maximum symmetric strategy for $EU$ implies that all infinitesimal symmetric transformations of $g$ cannot increase expected utility. Thus if $g+\delta$ is a strategy,

$$(D(EU))_{<g>}(<\delta>) \leq 0. \tag{3}$$

Now consider what happens if the $i$-th agent replaces their strategy $g$ with another strategy $g+\delta$. Equation (2) implies that the change in expected utility is

$$EU(<g>+\delta^i) - EU(<g>) = (D(EU))_{<g>}(\delta^i).$$

By the previous lemma, this is

$$EU(<g>+\delta^i) - EU(<g>) = (D(EU))_{<g>}(<\delta/n>).$$

The set of strategies is convex: so if $g+\delta$ is a strategy, then so must be $g+\delta/n$. Hence, since $g$ is a local maximum symmetric strategy for $EU$, the above difference must be less than or equal to zero. Hence the agent $i$ can't increase expected utility by unilaterally changing their personal strategy. Since all agents are identical and pursuing identical strategies, $<g>$ must be a Nash equilibrium. □

Let us call $o \in \mathbb{O}$ a definite observation for $g$ if $g(o)$ is either 0 nor 1. Thus definite observations are observations that entirely determine the agent's actions, without any randomness. Then the above proof leads to the following interesting corollary:

**Corollary 2.3.** *If $g$ is a local maximum symmetric strategy for EU and if $g$ has a non-definite observation, then the Nash equilibrium at $<g>$ is a non-strict equilibrium (i.e. a single agent can change their strategy in a way that doesn't make themselves worse off).*

*In fact, a single agent can change the probabilities of all their non-definite observations, without changing the expected utility.*

*Proof.* Choose a $\delta$ such that $g+\delta$ and $g-\delta$ are both strategies. This implies that $\delta(o) = 0$ whenever $o$ is an definite observation (or else one of $(g+\delta)(o)$ or $(g-\delta)(o)$ be below 0 or above 1, and hence would not be a probability). It can also be seen if $o$ is a non-definite observation for $g$, then a $\delta$ with the above properties and $\delta(o) \neq 0$ can be found (if a probability is neither 0 nor 1, it can be moved at least a little bit in both direction).

Since $g$ is a local maximum symmetric strategy for $EU$, equation (3) implies both $(D(EU))_{<g>}(<\delta>) \leq 0$ and $(D(EU))_{<g>}(<-\delta>) \leq 0$. This is only possible if $(D(EU))_{<g>}(<\delta>) = 0$. Hence

$$\begin{aligned} EU(<g> + (\delta)^i) &= EU(<g>) + (D(EU))_{<g>} <\delta/n> \\ &= EU(<g>) + (1/n)(D(EU))_{<g>} <\delta> \\ &= EU(<g>). \end{aligned}$$

Thus a single agent can change their strategy by $\delta$ without changing expected utility: the Nash equilibrium is weak.

Conversely, assume agent $i$ has changed her strategy to $g + \delta$, such that $\delta(o) = 0$ on any $o$ that is a definite observation of $g$'s. Then there exists a $\mu > 0$ such that $g + \mu\delta$ and $g - \mu\delta$ are both strategies. Hence

$$(D(EU))_{<g>} <\mu\delta> = 0$$

which, since the derivative is linear, means that $(D(EU))_{<g>} <\delta/n> = 0$ and hence that

$$EU(<g> + \delta^i) = EU(<g>).$$

Consequently agent $i$ changing their strategy to $g+\delta$ has left the expected utility unchanged. A single agent can thus change the probability of implementation for any non-definite observations. □

# References


[BDS12] Nick Bostrom, Thomas Douglas, and Anders Sandberg. The unilateralists curse: The case for a principle of conformity. *in preparation*, 2012.