

Presentation at FMEC, Oslo, Norway, June 11-12, 2001:

On a Dual Interpretation of Evolutionary Dynamics*

by

Steven O. Kimbrough

University of Pennsylvania

3620 Locust Walk, Suite 1300

Philadelphia, PA 19104-6366

kimbrough@wharton.upenn.edu

<http://grace.wharton.upenn.edu/~sok/>

*File: oslo-evolutionary-duality-slides.tex.

File: oslo-evolutionary-duality-slides.tex. Created: 2001-6-2. Modified: 2001-6-7.

See especially: evolutionary-duality.tex in my agebook.tex manuscript.

/ Header */*

Outline

- Context: Design of intelligent agents (& e-commerce)
- Goals and motivation
- Evolutionary game theory & strategy-centric agents
- Identity-centric agents
- Dual interpretation of the evolutionary dynamics
- Darwinian theories of mind
- Naturalizing intentionality: mental states & natural selection
- So what? (discussion)

Context

- Problem: How to design and build intelligent agents for e-commerce (broadly taken)?
- Required: (1) adaptive play in (2) strategic situations
(Or is it a requirement? See later.)
- “AI has solved all these problems. The techniques just have to be applied.”
(1) Not! (2) “just” is unjust.
- Throughout: attend to computation, to algorithms
- Begin: look at evolutionary game theory.
Game theory: classical, behavioral, evolutionary.

From *How Brains Think* by William Calvin, pp. 50-1

Social life, besides facilitating the spread of new techniques, is also full of interpersonal problems to be solved, like pecking orders. You may need to hide food from the view of the dominant animal, in order to keep it for yourself. You need a lot of sensory templates to avoid confusing one individual with another, and a lot of memory to keep track of your past interactions with each of your colleagues. The challenges of social life go well beyond the usual environmental challenges to survive and reproduce that the solitary orang confronts.

So, strategic interactions—games—bid fair as contexts for study.

Goals and motivation (for this talk)

- Among friends, a PBI. Also: a proto-position statement.
- We want to build intelligent agents who will live and act in strategic situations. But we have a nagging (philosophical) puzzle to deal with: intentionality and the mind.
Are these connected and if so, how?
- Regardless of applications, we'd like a theory—preferably a naturalization—of intentionality.
- For applications, we (should) want to know why and how intentionality—seemingly so central to human mentality—is or is not relevant to actually building and fielding artificial agents.

Evolutionary game theory & strategy-centric agents

The briefest possible summary might be put as follows:

1. The study of games has divided into classical game theory, behavioral game theory, and (more recently) evolutionary game theory. All three approaches seek to predict and explain outcomes of strategic situations.
2. Classical game theory has produced a series of results that place rather severe, and disappointing, limits on our understanding of strategic situations, and our concept of rationality.
3. Evolutionary game theory, taking a quite different tack and essentially ignoring considerations of rationality, has produced a series of results that in many ways complement classical game theory and provide a more complete understanding of games.
4. The main technique of evolutionary game theory has been investigation of replicator dynamics (in a broad sense). Replicators in a Darwinian population are identified with game strategies and the dynamics of the population's composition are studied. Agents and strategies are essentially identical.

5. We propose to investigate the dynamics of populations that include more complex and intelligent agents, agents that may themselves have dynamics of learning and adaptation. Such investigations are required if certain important and now open questions in game theory are to be answered.

Evolutionary game theory & strategy-centric agents

- evolutionary game theory: strategy-centric agents

Many, very interesting results, but we want more.

- Needed: identity-centric (game-playing) agents

How do you do it?

Outline

- Context: Design of intelligent agents (& e-commerce)
- Goals and motivation
- Evolutionary game theory & strategy-centric agents
- Identity-centric agents

⇒ Dual interpretation of the evolutionary dynamics

- Darwinian theories of mind
- Naturalizing intentionality: mental states & natural selection
- So what? (discussion)

Dual interpretation of the evolutionary dynamics

- The simple GA
- Evolutionary dynamics.
Strategy-centric agents. Impressive findings of.
- Dual interpretation of the evolutionary dynamics.
Discussion.

Simple Random Evolution Regime Part 1 of 2: Initialize

1. Set populationSize (e.g., to 1000).
2. stopCondition $\leftarrow \perp$ (Set the stopping condition to false.)
3. Declare \mathcal{S} the set of possible replicators/strategies.
4. Declare currentPopulation.
5. Create the initial population, of size populationSize, using $s_i \in \mathcal{S}$. (According to your policy, possibly with randomness.) Place this population in currentPopulation.

6. Declare $\text{currentScore}(s)$ for each $s_i \in \mathcal{S}$ and in currentPopulation .
7. $\text{currentScore}(s) \leftarrow \mathbf{0}, \forall s$
8. Set $\text{numGamesPerGeneration}$ (number of games played each generation, e.g., 500).
9. $\text{numGamesSoFar} \leftarrow \mathbf{0}$
10. $\text{generationsSoFar} \leftarrow \mathbf{0}$
11. Set maxGens (e.g., to 100).

Simple Random Evolution Regime, Part 2 of 2: Run

Run until stopCondition is \top (true):

1. Do while numGamesSoFar < numGamesPerGeneration:
 - (a) Randomly pick two strategies/replicators, s and t , from currentPopulation.
 - (b) Play s and t against each other and record their scores, score1 and score2:
 - i. $\text{currentScore}(s) \leftarrow \text{currentScore}(s) + \text{score1}$
 - ii. $\text{currentScore}(t) \leftarrow \text{currentScore}(t) + \text{score2}$
 - (c) numGamesSoFar++ (increment by 1)

2. $\text{generationsSoFar}++$
3. If $\text{generationsSoFar} > \text{maxGens}$ then $\text{stopCondition} \leftarrow \top$.
4. $\text{numGamesSoFar} \leftarrow 0$.
5. Create the next generation, in currentPopulation in proportion to replicator/strategy strength currently measured by currentScore .
6. $\text{currentScore}(s) \leftarrow 0, \forall s$

Dual interpretation

Suppose, instead of the simple random regime, we had two populations. Instead of currentPopulation we have population(1) and population(2) and we make the obvious adjustments to the algorithm. What do we have then?

Now when we play two replicators against each other and get their scores the players are from different populations. Even if initialized identically, random drawing of replicators/strategies will likely lead to divergence in the constitutions of the two populations, as the generations ensue.

Observation:

- It is appropriate, and useful, to consider each of our two populations as identity-oriented agents.

Points arising:

1. We might call such an agent a simple, random, pool (SRP) agent. Such an agent, as well as those in a large class of relatives, is really a management mechanism for a partitioned off evolutionary system. Strategies are to these agents like cells in your body: individual cells are expendable, they come and go. A very different picture of the constitution of an agent; contrasting with agents-as-data-structures-plus-smart-algorithms views.
2. What are the interesting variations on this SRP agent? There are many. We'll discuss just a few.
3. First variation: continuous generations. The idea is to keep a continual updating of the scores on the agents. They can be initialized arbitrarily, but

after that, they are simply updated after each episode (e.g., play of the game) and never zeroed out (as in step (6)). We just keep a running tally.

Note: (a) This is an inherently more efficient approach. We don't need a large population for a small number of strategies. We don't need a population any larger than the number of strategy/replicator types in it. (b) Continuous generations will be more rapidly responsive to information obtained in play.

4. There is a strong connection between our SRP agent with continuous generations and classifier systems. In fact, the former is a special case of the latter. We should think even more generally: classifier systems are examples of encapsulated evolutionary (Darwinian) processes (EDP). The encapsulation—via a mechanism to manage the pool of replicators—creates something we can identify as an individual, having an identity that persists over time and change. We might contrast this with open or naked Darwinian processes (NDP).

5. Second variation: perception and/of state. This is part of classifier systems as originally conceived by John Holland. The idea is that the environment presents the agent with state information (recall our reinforcement learning framework in the previous chapter). Computationally, this could be presented as a bit string of 1s and 0s. The agent's encapsulation mechanism would then pick a replicator to "fire." Typically done by "matching" and then strength-biased selection.

Note: The NDP analog is a soup of replicators, but not everyone can play everyone, in the circumstance.

6. Third variation: too many states. The number of possible states may be so large that if we enumerate them and try to learn about them all at once, then learning will be much too slow. Categorization is necessary; avoidance of unnecessary detail. Or the number of possible states may be simply too large

to even consider expressing explicitly in an agent. What to do? Categorization and the fourth variation.

7. Fourth variation: composing responses. Koza and ADFs. Classifier systems: classifiers as condition-action pairs that can “call” other classifiers.

Enough of this (for now). The dual interpretation of evolutionary processes—strategy-oriented:identity-oriented::NDP:EDP—is useful and fruitful.

Outline

- Context: Design of intelligent agents (& e-commerce)
- Goals and motivation
- Evolutionary game theory & strategy-centric agents
- Identity-centric agents
- Dual interpretation of the evolutionary dynamics

⇒ Darwinian theories of mind

- Naturalizing intentionality: mental states & natural selection
- So what? (discussion)

Darwinian theories of mind

Darwinian system (engine): “Descent with modification by natural selection.”
Variation and selective retention (with copying).

- Replicators vary, can be copied (imperfectly), and are subject (directly or indirectly) to selection.
- Variation is blind. Random or not, it is non-Lamarckian.
- Selection may operate directly on the replicators (genotypes), indirectly on the phenotypes, and even indirectly on the group.

Emphasize: variation comes first, then selection.

Darwinian theories of mind

- evolution producing the mind/brain vs evolutionary operation of the mind/brain

A remarkable parallel, which to my knoweldge has never been noticed, obtains between the facts of social evolution and the mental growth of the race, on the one hand, and of zoölogical evolution, as expounded by Mr. Darwin, on the other.

Edelman, Calvin, ...

William James, *The Atlantic Monthly*, “Great Men, Great Thoughts, and the Environment,” vol. XLVI, October 1880, no. CCLXXVI, pages. 441–459.

Transition

- Evolutionary dynamics: a great, but limited way to investigate populations of agents in strategic contexts. Need an identity-centric view.
- This dual interpretation basically internalizes the evolutionary dynamics, which we can now see as a very special case of reinforcement learning.
- Picture is this: ideas/strategies evolve in the head by variation and selective retention. May or may not involve “genetic” operations.
- But we have a nagging puzzle to deal with: intentionality and the mind.

Outline

- Context: Design of intelligent agents (& e-commerce)
 - Goals and motivation
 - Evolutionary game theory & strategy-centric agents
 - Identity-centric agents
 - Dual interpretation of the evolutionary dynamics
 - Darwinian theories of mind
- ⇒ Naturalizing intentionality: mental states & natural selection
- So what? (discussion)

Sentences with propositional content

Examples (with content P): S believes that P, S promises that P, S says that P, etc.

- Saying that (indirect discourse):
 - Galileo said that the earth moves
 - Bush said that the U.S. economy was in deep trouble and only massive tax reductions, especially for the wealthy, could save the day.

- Propositional attitude descriptions (aka: clausal complements) such as:
 - Jane believes that Tom loves Mary
 - Jane desires that Tom does not love Mary
 - Jane intends that Tom will marry Susan

and their stylistic equivalents. For example, Tom intends to marry Mary is arguably a variant of the more stilted Tom intends that he (Tom) will marry Mary.

- Speech act descriptions, such as:
 - Tom promised that he will marry Mary.

- Sam promised Sue a diamond.*
- Bush asserted that his administration would operate on a bipartisan basis.
- The Supreme Court declared that the ballots will not be counted.

- Modal descriptions, such as:
 - It is impossible that Gore can appeal a Supreme Court decision.
 - Necessarily, it is raining or not

- Deontic descriptions, such as:

*I take this kind of construction as shorthand for, here, something akin to Sam promised Sue that she (Sue) will get a diamond.

- Jake is obliged to Tom that Jake read(s) the letter.
- Parking is not permitted here.

- Perceiving-that descriptions, such as:
 - Jane saw that Tom kissed Mary
(as distinguished from Jane saw Tom kiss Mary.)

- Others, e.g.,
 - Gore needs it to be the case that a miracle occurs.

Comments

1. Most of these examples and other cases involve some degree of inten[st]ionality.

Above: focus on intensionality (“s-ality”). Roughly: linguistic description of intentionality.

Diagnostic: failure of substitution of equivalents (identicals) can fail to preserve truth. “S believes that P, P if and only if Q, therefore S believes that Q” is surely invalid, as is “S knows that a robbed the bank, $a = b$, therefore S knows that b robbed the bank.” Aside: degrees or levels of inten[st]ionality.

2. Here: focus on intentionality, intentional states

More on inten[st]ionality

Puzzling. Thought to challenge naturalism, materialism. The “mark of the mental”? Intentionality: directed, aspectual.

$$f(x) = z, x = y \models f(y) = z$$

$$P, P \leftrightarrow Q \models Q$$

$$P \rightarrow R, P \leftrightarrow Q \models Q \rightarrow R$$

Examples of intentionality: Oedipus, Juliet, police knowing the bank robber, BDI, etc.

Levels of inten[st]ionality, include: equivalences: extensional (Batman and Bruce Wayne), nomic (laws of nature), alethic (ordinary necessity), analytic (triangles, bachelors), synonymy (?)

Modal approaches (to intensionality)

Standardly:

Modal logic, e.g., $\Box P, P \leftrightarrow Q \not\models \Box Q$ but $\Box P, \Box(P \leftrightarrow Q) \models \Box Q$

Logics of belief, of knowledge, etc.

Speech act analysis pushed by sok:

$$\text{promise}(P) \approx \exists e(\text{promise}(e) \wedge \Box(K(e) \leftrightarrow P))$$

What's wrong with this? (a) Does there have to be something wrong to consider an alternative? (b) In my own case, I've been bothered by the lack of principled motivation for a modal approach. It's there for intensionality (only?).

Aside: Disquotation theory/approach: core idea

Propositional content has (at least) two important aspects. First, it is about something, that is to say it is true-or-false or rather it is a description, accurate or not, of something. Second, it is itself something about which we attribute certain properties, e.g., that Mary believes it or hopes it or asserts it or promises it. Summarizing (perhaps sloganizing), we might put the point by saying that the sentences of interest here large have the structure: content + comment (on the content). The core idea I wish to develop involves directly recognizing and representing these two aspects (content, comment) of sentences with propositional content.

Here: Intentionality

- Mark of the mental? Puzzling.
- Where does intentionality come from?
- How can an intentional state arise? Arise naturally, not “from us”?

Meaning and intentionality

- Grice: natural and non-natural meaning
- Dretske on information, which is just improbability. Ask what is $\Pr(F|i)$? If it is 1, then i indicates F , carries the information that F .
- Emphasize: indication/information is an objective relationship and is entirely distinct from what anyone knows or believes about the relationship.

Indicators

How do you get to be an indicator? Well, suppose i indicates F (indicator indicates fact). How can i achieve this status? Several ways, including.

1. F causes i . Smoke means/indicates (Grice and natural meaning) fire because smoke causes fire.

Note: weaker types of association will do, too, but this need not detain us.

2. Causal antecedants. G causes F and i . What I see on my TV screen indicates what you see on yours, given that we are watching the same show. Channel conditions.

3. Guessing. We tend to think of i as a signal, given off by, caused by, F , but that is only one way i can come to indicate F . The Sixers are playing tonight. Either they will win or they will lose. Flip a coin. Suppose heads has come up. If the Sixers win, then heads has indicated that; if the Sixers lose, then heads has indicated that.

Sounds funny, odd, but this is just a consequence of unpacking the concept of indication. What's odd is that guessing is so useless to us but indication is, or seems, so helpful.

Dretske: “If You Can’t Build One, Then You Don’t Know How It Works”

Compass needles are about geographical regions or directions under one aspect (as, say, the direction of the pole) and not other (as, say, the habitat of polar bears). This is the same way our thoughts are about a place under one aspect (as where I was born) but not another (as where you were born). If having this kind of profile is, indeed, one thing that is meant by speaking of a state, condition, or activity as intentional, then it seems clear that there is no need to naturalize intentionality. It is already a familiar part of our physical world. It exists wherever you find clouds, smoke, tree rings, shadows, tracks, light, sound, pressure, and countless other natural phenomena that carry information about how other parts of the world are arranged and constituted.

Intentional systems, then are not the problem. They can be picked up for a few dollars at your local hardware store.

Dretske: “If You Can’t Build One, Then You Don’t Know How It Works”

The problem is the supplier of information has to have the function of supplying information. How does that work?

A supplier of information acquires the function of supplying information by being recruited for control duties because it supplies it.

Then, page 223 (quote)!

Dretske: “If You Can’t Build One, Then You Don’t Know How It Works”

Page 223, recipe for a “proto-thought” Take a system that has a need for the information that F , a system whose survival or well-being depends on its doing A in conditions F . Add an element, or a detector system that produces elements, that carries information about condition F . Now, stir in a natural process, one capable of conferring on the F -indicator the function of carrying this piece of information. One does not quite “stir” these processes in (the metaphor is getting a bit strained at this point). Once you have got the right system, adding functions is more like waiting for the dough to rise. There is nothing more one can do. You sit back and hope that natural processes will take a favorable turn. . . . If all goes well, when the process is complete, the result will be a system with internal resources for representing—and, equally important from the point of view of modeling the mind, misrepresenting—its surroundings. Furthermore, that this system represents, as well as what it represents, will be independent of what we know or believe about it. For we, the cooks, are not essential parts of this process. The entire process can happen “spontaneously” and, when it does, the system will have its own cache of original intentionality.

Hmmm... a bit vague, but:

- natural selection is intentional!

Selection vs selection for.

- natural selection induces intentional states in the replicators
- this holds for both NDPs and EDPs
- Dretske's "recruitment" now has a natural operationalization. Many, actually. Think of Koza's ADFs, &c.!

Outline

- Context: Design of intelligent agents (& e-commerce)
 - Goals and motivation
 - Evolutionary game theory & strategy-centric agents
 - Identity-centric agents
 - Dual interpretation of the evolutionary dynamics
 - Darwinian theories of mind
 - Naturalizing intentionality: mental states & natural selection
- ⇒ So what? (discussion)

So what (in regard to practice)?

It's theoretically and philosophically interesting if we can build a mind, even a proto-mind. Why should this matter for applications? Answer: Theory can guide us, can give us a general and suggestive map, can set expectations, can help us frame problems. Here, I want to tell a story; goes like this.

Agents use representations (putative indicators) to condition on when learning and deciding what to do. These representations may be of three kinds: built-in (innate) assumptions, sensor data, and memories. Agents also face a meta-problem of whether and how to use the representations (think of feature extraction in a digitized image). Rather, this problem interacts with problem of what to do given the indications at hand. A natural approach is to organize the representations into modules or building blocks (and categories within). So we have: (a) representations and manipulations thereof, (b) action rules, and (c) a reward structure/system that directly or indirectly sets goals.

Adaptability comes in degrees and may be multidimensional (speed and scope, e.g.). Different learning mechanisms (algorithms) have different adaptability characteristics. What is optimal is in general a strategic matter and will depend on context (especially what the other agents are doing). Zero adaptability (as distinguished from changeability, maintainability) is most efficient only in a very narrow sense: no waste, no mistakes, every action optimal, but neither are there any new discoveries. So when would we want our agents to make discoveries and how much waste

or error (non-optimal behavior) are we willing to tolerate? When would we encourage learning on manipulations of representations? On action rules? On intermediate goals?

So the picture is this. Agents need (a) representations, (b) decision/action rules, and (c) goals. If you fix (a) and (b), you don't need to worry about (c). Example: classical rule-based systems. Fix (a) and (c), and float (learn on) (b), you get standard reinforcement learning, including the dual (identity-centric) interpretation of the evolutionary dynamics. Float (a) and you add a layer of complexity and you get intentionality à la Dretske. Float (c) and you get more complexity and flexibility.

Could we be more specific?

- For offline discovery, e.g., Tit-for-tat. Even very complex, nonlinear situations.
- For nonstationary, dynamic processes.
- When it's required for an activity that is worthwhile: strategic and indeterminate situations (repeated games). Ecological/social learning. Point of the philosophical excursion: misrepresentation is essential to mentality (perhaps); not just because things are unreliable, but because indication—information—creates intentional states and the mechanisms of mentality exacerbate this. Misrepresentation is the price of intelligence.
- When it is too expensive to implement with less adaptability.

A little more specific?

[Discuss]

Get: William Calvin, “The Brain as a Darwin Machine,” *Nature*, 330, 33-44, 1987.

William James, *Atlantic Monthly*, 1880.

Note: This talk is something of a position paper in David Wood’s sense.

Turning practice into theory.

Beyond para. 1 of Calvin, saying “Intelligence is for when you don’t know what to do.” Contrasted with smarts. We can now give a more nuanced account than this. Intelligence is about what kind of risks you’re willing to take. The more ventured, the more potentially gained, and lost (including time to learn). We need to investigate the ecology of intelligence, e-commerce offers a fine venue. I think the main philosophical puzzles about representation and intentionality are well handled. (Consciousness? Stay tuned.) Well-handled enough that the road to further scientific progress is freshly cleared and ready for transversal.

`/* $Header$ */`