Draft of March 27, 2008; comments welcome

Interpretive Economy, Schelling Points, and evolutionary stability*

Christopher Potts
UMass Amherst

Abstract I attempt to show that Kennedy's (2007) Interpretive Economy principle follows from basic assumptions about cognitive prominence and evolutionary stability.

Keywords: comparatives, economy, Schelling points, evolutionary game theory

1 Introduction

Kennedy's (2007) Interpretive Economy principle demands that speakers make maximal use of the conventional meanings of the words and phrases they encounter, resorting to context-dependent values only where linguistic convention ends. Kennedy calls on the principle mainly to explain a limitation on the interpretive variability of non-comparative uses of scalar adjectives (section 2 below). My goal for the present short paper is to show that Interpretive Economy follows from basic assumptions about cognitive prominence and evolutionary stability. Thus, we can leave the principle out of the semantics and the pragmatics but still have a precise account of the effects it was meant to determine.

2 Scalar adjectives and Kennedy's puzzle

Kennedy & McNally (2005) and Kennedy (2007) develop a typology of gradable adjectives based on the scales with which they conventionally associate. The following summarizes with representative examples:

* My thanks to David Beaver, Robin Clark, Chris Kennedy, the participants at the LSA Institute Workshop *Conversational Games and Strategic Inference*, Stanford, July 11, 2007, and the participants in Ling 753, UMass Amherst, Spring 2008. None of these people necessarily endorses the proposal herein.

(1)

totally open	○——○	*short*, *expensive*
lower closed	●——○	*wet*, *bent*
upper closed	○——●	*pure*, *straight*
totally closed	●——●	*opaque*, *open*

A filled circle indicates a scalar endpoint. An open circle indicates unboundedness. An adjective's scale figures centrally in judgments about truth, and it affects patterns of adverbial modification and interpretive variability.

For example, if something is even slightly bent, then it counts as bent (setting pragmatic slack aside), but there is no upper limit on the degree to which something can be bent, making ??*completely bent* quite strange. The reverse holds for *straight*: absolute straightness sets the upper-bound (and thus *completely straight* is well formed), and an object can deviate arbitrarily far from that standard (hence, ??*slightly straight*)

With closed-scale adjectives like *open*, both the maximal and the minimal endpoints are prominent (*slightly open*, *completely open*). If we want to talk privately, I might first ask whether the door is open. In this case, even the minimal degree of openness counts as open. In contrast, if our goal is to get a large couch through the doorway, then the question of whether the door is open is likely to be about whether it has the maximal degree of openness.

Adjectives totally open scales provide no endpoints to rely on (??*slightly tall*, ??*completely tall*), so our evaluations are based on contextual standards. These can be highly variable. For example, the standard for *tall* depends on what we're talking about (chipmunks, people, giraffes) and where and when we're talking about it. There is clearly no maximal degree of tallness. The claim that there is no minimal degree is harder to motivate, but it follows from the assumption that *tall* and *short* share the same scale and the observation that there is no maximal standard for shortness. (??*completely short*; Kennedy 2007: 34–35 addresses this in more detail.)

The relevance of endpoints is nowhere more evident than in simple predications involving these adjectives. For adjectives with totally open scales, our judgments depend on where we set the contextual standard. This standard is the main factor in whether we judge sentences like (2) true, and it figures centrally in how we determine the referents for definite nominals like those in (3).

(2)
a. That cup is tall.
b. That bicycle is expensive.

(3) a. the tall cup
b. the expensive bicycle

The situation is different for adjectives with partly or completely closed scales. For them, only the endpoints are available as standards for interpretation. For example, (4a) is true just in case the bar is totally straight, (4b) is true just in case the towel has even the smallest amount of moisture on it, and (4c) is ambiguous between minimal and maximal interpretations, rather than being vague about where the standard is set.

(4) a. The bar is straight.
b. The towel is wet.
c. The door is closed.

We might, on occasion, speak a little loosely, in the sense of Lasersohn 1999, but, strictly speaking, truth and the notions defined in terms of it are determined by the endpoints wherever they exist. There is no possibility in these cases of fixing an arbitrary contextual standard as in (2)–(3). Kennedy (2007) reviews a great deal of evidence for this conclusion. Here is a brief summary:

(5) a. Adult experimental subjects presented with two jars, neither full, disprefer utterances like *Please hand me the full one*, whereas corresponding utterances like *Please hand me the tall one* are fine even when neither object is particularly tall in an absolute sense (Syrett, Bradley, Kennedy et al. 2005).
b. The Sorites Paradox is unattested for absolute adjectives, but it is precisely the contextual standard that gets us into the paradox in the first place — one wonders why we can't force it with absolute adjectives by appeal to a relative interpretation.
c. We regard examples like *My hands are not wet, but there is some water on them* as inconsistent, though a relative analysis of *wet* would allow consistent readings where the first conjunct said simply that the speaker's hands were below the contextual standard for wetness (Kennedy & McNally 2005: 359).

One might respond to all this by saying that the typology in (1) is mistaken in that only totally open scales are genuinely accessible to the grammar, whereas the others restrict access to the endpoints. However, this semantics wrongly predicts that these adjectives can't appear in comparatives, where any point might turn out to be interpretively relevant:

(6) a. This mug is fuller than that one.

b. This rag is less wet than that one.

c. My office door is more open than I would like.

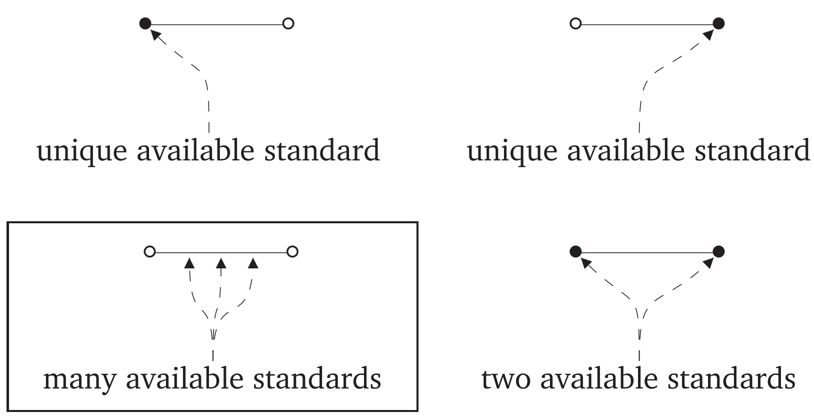
Similarly, it leaves unexplained the fact that these adjectives can be modified by non-scalar-endpoint adverbials (e.g., *extremely wet*, *a little bit open*). There seems no avoiding the fact that the full, rich array of points in (1) is accessible by grammatical means — just not in simple examples like (4).

Thus, we have arrived at the puzzle Kennedy confronts:

(7) **Kennedy's puzzle** Why can't closed-scale adjectives be interpreted relative to arbitrary contextual standards? They have the scale structure to support it, and the language provides the relevant mechanisms, as evidenced by the meanings of adjectives with totally open-scales.

The following picture summarizes the puzzle:

(8)



3 Interpretive Economy and its drawbacks

Kennedy's (2007: §4.3) seeks to solve (7) by imposing Interpretive Economy:

(9) **Interpretive Economy** Maximize the contribution of the conventional meanings of the elements of a sentence to the computation of its truth conditions.

The idea is that endpoints are conventional standards. Interpretive Economy therefore demands that we make use of them if they are present in the underlying scale structure. The principle makes intuitive sense. Why resort to the instability of free variables and the like if you can depend on fixed conventions? Kennedy writes:

> although participants in a discourse may not be in full agreement about those properties of the context that play a role in the computation of context-dependent features of meaning, they are in agreement about the conventional meaning of the words and complex expressions in the sentences they use to communicate (Kennedy 2007: 36).

We need not assume that the picture is as neat as this suggests, with linguistic conventions absolutely reliable and context dependency always a risk. As long as the conventional meanings are relatively less variable than the context-dependent ones, the strategy suggested by (9) makes sense.

However, though Interpretive Economy might be intuitively plausible, there is no getting around the fact that it is different in kind from the other parts of Kennedy's proposal. I see three central objections:

(10) a. It is the only aspect of the theory that is not grounded in functional denotations for morphemes or well-established forms of context dependency.

b. It sounds pragmatic, but it doesn't yield to pragmatic influences.

c. It is an optimization principle left unsupported by a theory of optimization (Jacobson 1997).

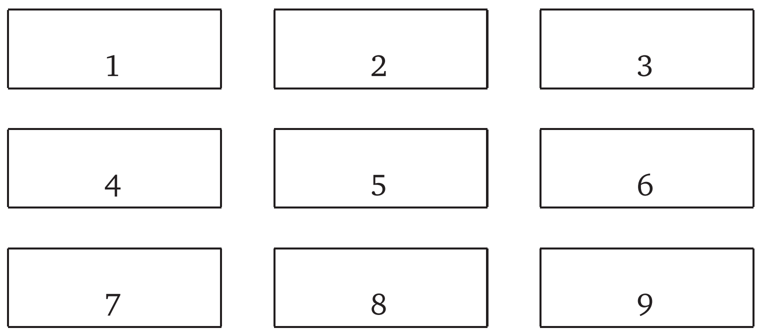
In addition, (9) seems to depend on the assumption that the endpoints are conventionalized meanings in some sense. But this is something we should explain in terms of more basic mechanisms.

Despite these drawbacks, I think (9) harbors important insights into the interplay between convention and context dependency. My goal, therefore, is to capture its effects, but in a way that avoids the conceptual challenges just mentioned. I seek to derive this pressure from two independently motivated principles of cognition: (i) scalar endpoints are cognitively prominent; and (ii) this prominence makes endpoint interpretations evolutionarily stable — to such a degree, in fact, that there is little reason to expect pragmatic flexibility and every reason to expect that they will behave like grammaticized features of these morphemes. (Whether they *are* grammaticized is a question I postpone until section 7.)

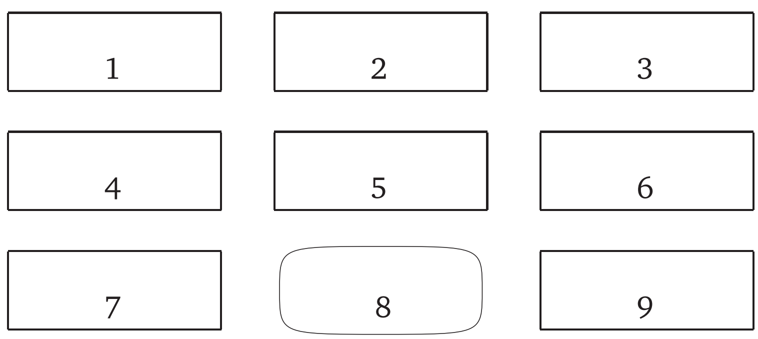
4 Schelling points and the typology of scales

Imagine that you've been asked to play the following two quick games with another player:

(11) **Game 1** Look at the following diagram and select a shape-token by number. Keep your choice secret! (It is crucial that you communicate nothing to each other.) You will be rewarded just in case you and your partner both select the same shape-token.



(12) **Game 2** Same as Game 1, but now with the following picture:



Which choices would you make? In general, few players of game (11) are rewarded. However, game (12) shows a high rate of return. Players overwhelmingly go for #8, the distinguished circle. It is what Schelling (1960) called a 'focal point', and what later became known as a *Schelling Point*. Schelling characterizes them as follows:

> Most situations — perhaps every situation for people who are practiced at this kind of game — provide some clue for coordinating behavior, some focal point for each person's expectation of what the other expects him to expect to be expected to do. Finding the key, or rather finding *a* key — any key that is mutually recognized as the key becomes *the* key — may depend on

> imagination more than logic; it may depend on analogy, precedent, accidental arrangement, symmetry, aesthetic or geometric configuration, casuistic reasoning, and who the parties are and what they know about each other (Schelling 1960: 57).

Clark (1996) contains additional discussion and is, to my knowledge, the first application of these ideas directly to linguistic phenomena, though they also figure in Lewis's (1969) theory of how people resolve linguistic coordination games with multiple equilibria.

No single principle determines the distribution of Schelling Points. You and I might succeed in the game based on (11) if we are both aware of the cognitive fact that, for newspaper readers, the upper left corner is the most salient. If we are number theorists or gamblers, we might pay attention to the numerical labels more than position or shape. And so forth. Lewis (1969: 35) writes that a Schelling Point "does not have to be uniquely *good*; indeed, it could be uniquely bad. It merely has to be unique in some way the subjects will notice, expect each other to notice, and so on."

People can find Schelling Points in abstract space as well. If players are presented with five numbers, four even and one odd, the lone odd number might emerge as a Schelling Point. It seems highly likely, then, that Schelling Points play a role in shaping language as well. I propose such a connection for the case of scalar adjectives:

(13) **Endpoints as Schelling Points** For purposes of scalar adjective interpretation, scalar endpoints are Schelling Points.

Simple, non-comparative gradable adjective predications present us with a range of interpretive choices. Our goal is to coordinate on a single shared meaning (any will do). This is a daunting task; without clues, we have little hope. It is therefore natural that speakers gravitate to prominent points where they are made available by the semantics. (The next section provides a formal interpretation of this idea in the context of strategic games.)

This conception is very much in keeping with the basic tenets of Kennedy's (2007) theory, which are directly tied to notions of cognitive prominence. Open-scale adjectives force us to choose a contextual standard, and Kennedy proposes that we do this in such a way as to "ensure that the objects that the positive form is true of 'stand out' in the context of utterance, relative to the kind of measurement that the adjective encodes" (p. 17). In confronting the more limited closed-scale interpretations, he calls even more directly on a Schelling-like assumption: "a maximal degree always stands out on an upper closed scale, and a non-minimal degree always stands out on a scale with a minimum (zero) element" (p. 36). In the case of closed-scale adjectives, both the maximal and minimal values stand out equally well, and both are, in turn, attested points of interpretation. Thus, (13) is implicit in Kennedy's (2007) proposal. I have merely brought it out and connected it with other aspects of cognition.

This alone will not suffice to solve Kennedy's puzzle, however. One could fairly object that this predicts relative readings of sentences like *The cup is full* when conditions are such that a non-maximal degree is salient. In (5a), for example, I briefly reported on experimental work by Syrett, Bradley, Kennedy et al. (2005) indicating that people are unwilling to accept utterances like *Please hand me the full one* when none of the salient containers is full (though some are fuller than others). Surely, though, one of the many degrees d that yields a well-defined referent for the definite (by placing exactly one object at least as high as d and all others below it) counts as prominent in this situation. Why not coordinate there?[1]

The next section attempts to answer this objection by showing that (13), properly implemented in the theory of games, determines evolutionary stability for interpretations at those endpoints. It seems reasonable to interpret this stability as the formation of a convention. Section 6 begins a conceptual assessment of the proposal.

5 The evolutionary stability of endpoint interpretations

The goal of this section is to show that endpoint interpretations, where available, are stable in a way that would result in an entire community gravitating towards them. To do this, I employ a simple theory of evolutionary dynamics, based on the discussion in Benz, Jäger & van Rooij 2005: §3 and informed by the linguistic analysis of Jäger (2007).

Let's begin with a simple scenario, based on one discussed by Benz, Jäger & van Rooij (2005: 50-52). The population consists of *A*s and *B*s. They always travel in pairs. The *A*s have an excellent sense of direction, so they never get lost. The *B*s have a terrible sense of direction. If a *B* is left without an *A* to guide it, then it gets lost, never to be heard from again. The situation can be depicted in matrix form:

(14)

	A	*B*
A	10	10
B	10	0

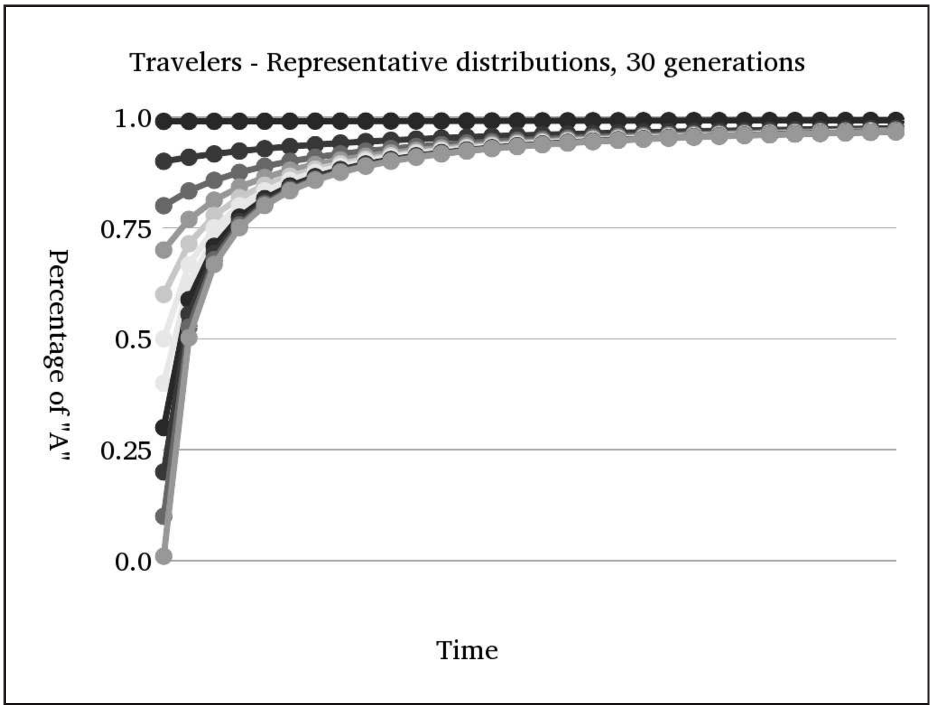
The matrix presents all possible pairs of *A*s and *B*s. I've given the payoffs (fitness

1 Indeed, Pasha Siraj suggests (p.c.) that this might be what people would do if forced to choose a referent irrespective of the acceptability of the instruction sentence.

measures) numerically, just for concreteness. The scenario requires only that the payoffs for $\langle A,A \rangle$, $\langle A,B \rangle$, and $\langle B,A \rangle$ be higher than those for $\langle B,B \rangle$.

What will happen if we begin with a single mixed population of *A*s and *B*s and allow them to travel around together? *A*s always thrive, no matter who they are paired with. Some *B*s thrive as well, because they get paired with *A*s. Other *B*s are not so lucky: they get paired with another *B* and end up hopelessly lost. Over time, the *B*s die out, and the *A*s take over. The following graph depicts this behavior for a representative sample of initial population distributions:

(15)



The vertical axis represents the percentage of *A*s in the population (from which we can calculate the percentage of the *B*s). The horizontal axis represents time. As the graph suggests, any mixed population of *A*s and *B*s will develop into a population containing only *A*s. The speed with which this happens depends on the initial population distribution and the numerical values of the payoffs in (14), but the result is always the same.

The replicator dynamics of Taylor & Jonker (1978) model this progression with calculations that balance fitness (the payoffs in (14)) with the percentage of each type of player in the environment. The next few definitions characterize one version of this process. I call the basic structures *symmetric evolutionary games*:

Definition 1. A symmetric evolutionary game is a structure (T_n, u), where

i. $T_n = \{1, \ldots, n\}$ is a set of nonoverlapping sets of entity-types, with the total population given by their union.

ii. u is a total function from $T \times T$ into real numbers, defining the *fitness* or *utility* of type pairs.

Matrices like (14) succintly represent these games.

Probability distributions model the proportions of entities of each type in the population at specific times:

Definition 2 (Population distributions)**.** Let $G = (T_n, u)$ be a symmetric evolutionary game. The percentage of type $i \in T_n$ is

$$P^t(i) = \frac{|T_i|}{|T_1 \cup \cdots \cup T_n|}$$

The fitness of a type i at time t is defined in terms of the percentage of things of type i (at t) and the utility function u (which is unchanging):

Definition 3 (Fitness)**.** Let $G = (T_n, u)$ be a symmetric evolutionary game. The fitness of type i at time t is

$$\tilde{u}_i^t = \sum_{j \in T} P^t(i) \cdot u(i, j)$$

The average fitness of the entire population at time t is a weighted average of the fitness of each type at t:

Definition 4 (Average fitness of the population)**.** Let $G = (T_n, u)$ be a symmetric evolutionary game. The average fitness at time t is

$$\tilde{u}^t = \sum_{i \in T} P^t(i) \cdot \tilde{u}_i^t$$

Finally, we come to the calculation that determines the dynamic process depicted in (15). There are many perspectives one can take on these dynamics (Cressmann 2003: §2 provides an overview). The definition I give involves updating the percentage of things of each type in the population, as in (15) above:

Definition 5 (Evolutionary dynamics)**.** Let $G = (T_n, u)$ be a symmetric evolutionary game. The following derives the population distribution at a time from the population distribution at the immediately preceding time and the fitness calculations at that time:

$$P^{t+1}(i) = \frac{P^t(i) \cdot \tilde{u}^t_i}{\tilde{u}^t}$$

To bring these definitions together, I return to the simple traveling game described above, fitting it into this formal framework. The game is $G = (\{A,B\},u)$, where u is defined by the matrix in (14). If we begin with $P^{t_0}(A) = \frac{1}{2}$, then we obtain the following initial measures of fitness:

(16) a.
$$\begin{aligned} \tilde{u}^{t_0}_A &= \textstyle\sum_{j \in T} P^t(A) \cdot u(A,j) \\ &= \big(P^{t_0}(A) \cdot u(A,A)\big) + \big(P^{t_0}(A) \cdot u(A,B)\big) \\ &= \big(\tfrac{1}{2} \cdot 10\big) + \big(\tfrac{1}{2} \cdot 10\big) \\ &= 10 \end{aligned}$$

b.
$$\begin{aligned} \tilde{u}^{t_0}_B &= \textstyle\sum_{j \in T} P^t(B) \cdot u(B,j) \\ &= \big(P^{t_0}(B) \cdot u(B,A)\big) + \big(P^{t_0}(B) \cdot u(B,B)\big) \\ &= \big(\tfrac{1}{2} \cdot 10\big) + \big(\tfrac{1}{2} \cdot 0\big) \\ &= 5 \end{aligned}$$

The average fitness of this population at t_0 is therefore

(17)
$$\begin{aligned} \tilde{u}^{t_0} &= \textstyle\sum_{i \in T} P^t(i) \cdot \tilde{u}^t_i \\ &= (\tfrac{1}{2} \cdot 10) + (\tfrac{1}{2} \cdot 5) \\ &= 7\tfrac{1}{2} \end{aligned}$$

With these calculations out of the way, we can calculate the percentage of As (and in turn the percentage of Bs, since there are just two types) at time t_1:

(18)
$$\begin{aligned} P^{t_1}(A) &= \frac{P^{t_0}(A) \cdot \tilde{u}^{t_0}_A}{\tilde{u}^{t_0}} \\ &= \frac{\frac{1}{2} \cdot 10}{7\frac{1}{2}} \\ &= \frac{2}{3} \end{aligned}$$

The graph in (15) above depicts 30 repetitions of this calculation.

With the technical background in place, let's now consider a different game: the population consists of people who study in the library and people who study in the department. Each member of the population is born to study in one place or the other. Studying in pairs is essential; people who study alone get bored, fall asleep, and flunk out. Members of this population are randomly paired up. The game depicted in (19) captures this situation.

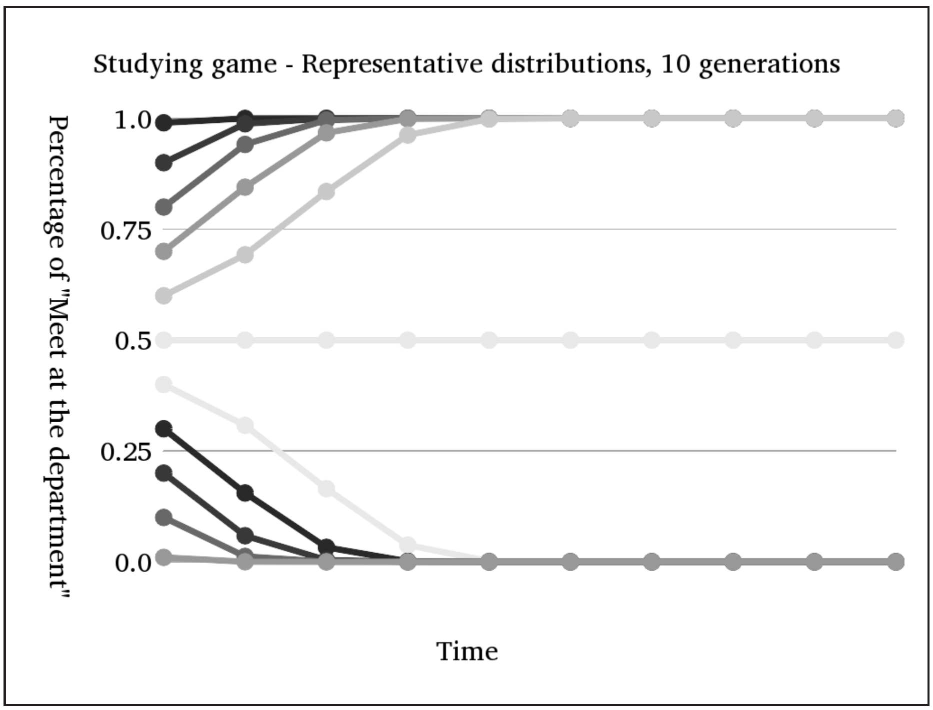
(19)

	Study in the dept	Study in the library
Study in the dept	10	0
Study in the library	0	10

This game rewards coordination. If the individuals in each pair are of the same type, they get together and excel. If they are of different type, they study alone, fall asleep, and flunk out. As in the traveling game (14), we just require coordinating to be strictly better than not coordinating. The difference can be tiny.

Definition (5) determines the following for various representative distributions:

(20)



Where the population is evenly divided, it remains stable. Where there is an imbalance, however slight, in the initial state, the population eventually contains only members of that initial majority type; players of the other type fall asleep (die out). The speed with which this happens depends on how far apart the coordinating and noncoordinating payoffs are, but the outcome is always the same in the limit.

My claim is that the game of scalar adjective interpretation has the same basic form as the studying game, and thus the evolutionary dynamic is basically that of (20). The relevant notion of coordination in this linguistic game is that the speaker's intended message is the same as the one that the hearer perceives; this is like meeting in the library or the department. Since no single adjectival meaning is reliably better than any other, the members of this discourse community don't care which convention they settle on, just as the choice of the library or the department is irrelevant. Finally, the Schelling Point assumption (13) is simply that the players are biased in favor of endpoint interpretations (where available). The diagram in (21) summarizes these assumptions with a schematic game structure. The types in this case are modes of interpreting the scalar adjective **full**. I use $[\![\textbf{full}_d]\!]$ to abbreviate roughly "interpreting *full* relative to degree d", with $\bullet$ representing the unique endpoint and all $d_i < \bullet$.

(21)

	$[\![\textbf{full}]\!]_\bullet$	$[\![\textbf{full}]\!]_{d_i}$	$[\![\textbf{full}]\!]_{d_j}$	$\cdots$
$[\![\textbf{full}]\!]_\bullet$	α	β_1	β_2	$\cdots$
$[\![\textbf{full}]\!]_{d_i}$	β_3	α	β_4	$\cdots$
$[\![\textbf{full}]\!]_{d_j}$	β_5	β_6	α	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

a. Communicative assumption: $\alpha > \beta_i$ for all i

b. Schelling assumption: $P^{t_0}([\![\textbf{full}]\!]_\bullet) > P^{t_0}([\![\textbf{full}]\!]_d)$ for all $d < \bullet$

If we restrict ourselves to just two types, then we can easily depict this graphically. In order to illustrate the flexibility of the two assumptions in (21), I exemplify with a game that implements those assumptions in a numerically very conservative fashion:

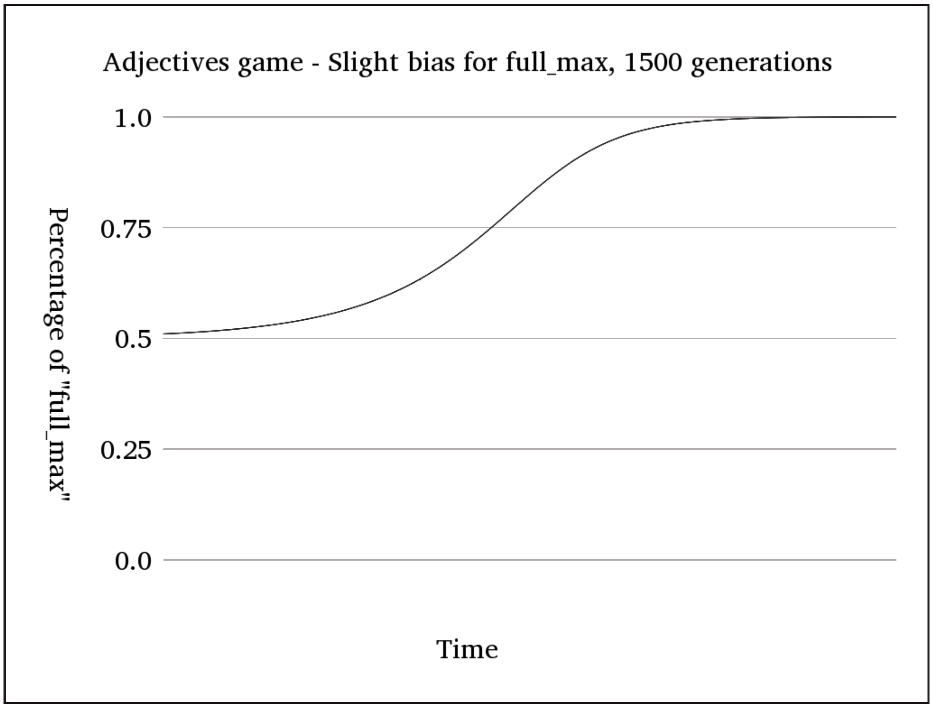
(22)

	$[\![\mathbf{full}]\!]_{\bullet}$	$[\![\mathbf{full}]\!]_{d}$
$[\![\mathbf{full}]\!]_{\bullet}$	10	9.9
$[\![\mathbf{full}]\!]_{d}$	9.9	10

Schelling assumption: $P^{t_0}([\![\mathbf{full}]\!]_{\bullet}) = 0.51$

Because the differences are small, the evolution is slow, but it eventually stabilizes in the requisite way:

(23)

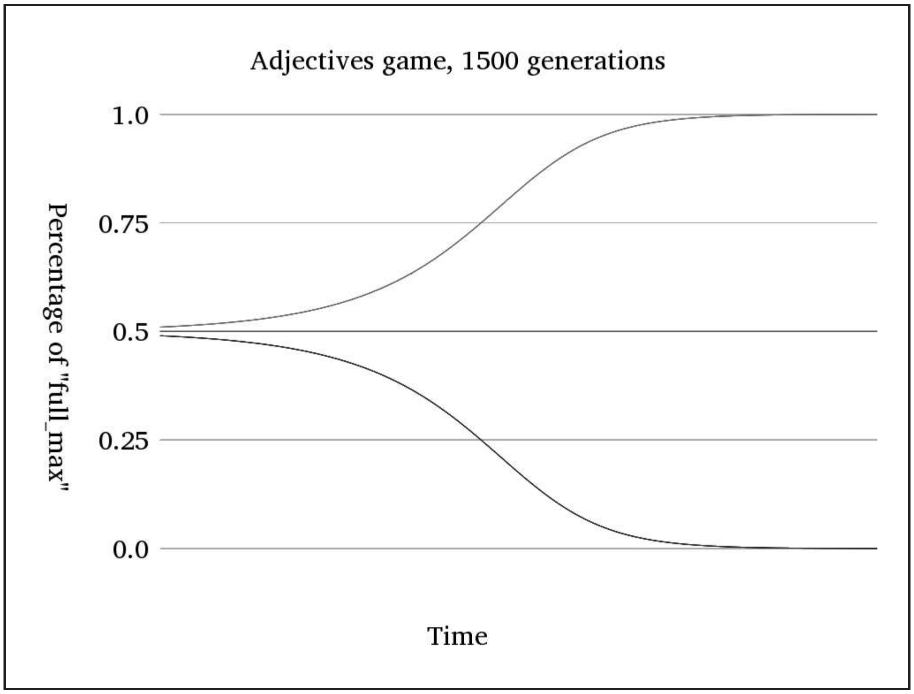


6 Assessing the stability

I advertised this proposal as one that can bring stability to certain interpretations. The relevant notion of stability is somewhat delicate, though. This is evident from comparing the graphs for the traveling game in (15) with the graph for the studying game in (20). In the traveling game, any mixed population leads to the same outcome; $\langle A,A\rangle$ is *evolutionarily stable* (Benz, Jäger & van Rooij 2005: 51). This is not true of the studying game, where the nature of the initial distribution greatly affects the outcome.

The comparative adjective game (22) is like the studying game in this regard. If the initial distribution favors $[\![\mathbf{full}]\!]_{d}$, then the entire population gravitates towards that interpretive scheme; if the initial population is evenly distributed, then that never changes:

(24)



I argue that this is as it should be. By assumption, the only thing that puts ⟦**full**⟧$_\bullet$ ahead of its competitors is that discourse participants are (perhaps ever so slightly) more likely to choose it. Just as in the Schelling Game (12), where players need not have any inherent preferences for one choice over another in order to win, so too do we succeed at the game of adjectival interpretation without inherently preferring endpoint interpretations for reasons of informativity, relevance, politeness, or anything else. We succeed because we have a suspicion about how others will play.

This flexibility receives factual support from scenarios in which new endpoints can be motivated by explicit communication among members of a speech community. For instance, imagine we work in a bar where the glasses are marked with a line indicating how much the bartender should fill them. That line furnishes a Schelling Point, and thus it might furnish what appears to be a relative (non-maximal) standard for fullness that can coexist with the more general absolute sense given by the glasses' rims. (Kennedy & McNally (2005: 371) discuss a related scenario.) In such cases, *full* might resemble *closed* in having two equally salient standards, even though only one is an endpoint. Similarly, if we are forming a basketball team and desire only tall players, then we might fix a single, explicit standard for tallness and then judge people as tall or not based on that fixed standard. Our calling attention to it in this way could make it a Schelling Point.

It is also significant that we require only a very slight bias for perfect communication (the diagonal pairs in the matrix depictions). With scalar adjectives, it might often be the case that many standards are good enough for current purposes. Certainly in the case of totally-open scales, it seems likely that we rarely, if ever, select exactly the same standard. Communication is successful as long as the difference in our choices doesn't matter. Payoffs might therefore be high even though we don't fully coordinate. However, as long as genuine coordination has the edge, however slight, it will eventually win out if we can figure out how to do it. Where present, scalar endpoints provide the requisite method: they are sufficiently salient to draw us to them reliably even if the gain in doing so is small.

7 Extra-grammatical explanation, simplified grammar

In sum, my proposal is that we have a slight bias for endpoint interpretations of scalar adjectives. The bias exists, not because endpoints are inherently superior, but rather simply because endpoints are (slightly) more cognitively salient than others. Using some of the apparatus of evolutionary game theory, I formalized this as a slight bias for endpoint interpretations, and the result of that initial-state assumption is a reliable evolution towards a population consisting entirely of such interpretations. I claim that this derives the intended effects of Kennedy's (2007) Interpretive Economy principle, and thus we can leave that principle out of his grammar of gradable adjectives. This leaves us with a theory that depends entirely on familiar context-dependent denotations.

Let me close with a question. Do the evolutionary dynamics predict grammaticization? A 'yes' answer seems reasonable. Over time, the endpoint interpretations become lexicalized, thereby leaving the realm of context-dependency and joining the realm of convention. Thus, we revise our denotations for *full*, *wet*, *open*, and the like. The revisions look stipulative, but we can point to their systematic origin. In this way, we can continue to capture the linguistic patterns without leaning directly on other aspects of cognitive life.

The alternative is to leave the lexical entries as they are. This means that individual gradable adjectives impose nothing special beyond their scale structure.[2] If we study them in isolation, we will appear to predict that simple closed-scale adjectival predications can be interpreted relative to contextually-supplied standards, and thus we will seem to run afoul of the data in section 2. But if we study these lexical items as part of the cognitive life of agents — agents who succeed at Schelling games like (12) — then we'll properly account for the data.

2 This in keeping with the view of Borer (2005a,b) that substantive lexical items are basically devoid of distinctive grammatical properties; see Potts 2008 for discussion.

References

Benz, Anton, Gerhard Jäger & Robert van Rooij. 2005. An introduction to game theory for linguists. In Anton Benz, Gerhard Jäger & Robert van Rooij (eds.) *Game Theory and Pragmatics*. Houndsmills, Basingstoke, Hampshire: Palgrave Macmillan.

Borer, Hagit. 2005a. *Structuring Sense*, vol. I: In Name Only. Oxford: Oxford University Press.

Borer, Hagit. 2005b. *Structuring Sense*, vol. II: The Normal Course of Events. Oxford: Oxford University Press.

Clark, Herbert H. 1996. *Using Language*. Cambridge: Cambridge University Press.

Cressmann, Ross. 2003. *Evolutionary Dynamics and Extensive Form Games*. Cambridge, MA: MIT Press.

Jacobson, Pauline. 1997. Where (if anywhere) is transderivationality located? In Peter Culicover & Louise McNally (eds.) *Syntax and Semantics*, vol. 29: The Limits of Syntax, 303–336. New York: Academic Press.

Jäger, Gerhard. 2007. Evolutionary game theory and typology: A case study. *Language* 83(1): 74–109.

Kennedy, Christopher. 2007. Vagueness and grammar: The semantics of relative and absolute gradable adjective. *Linguistics and Philosophy* 30(1): 1–45.

Kennedy, Christopher & Louise McNally. 2005. Scale structure and the semantic typology of gradable predicates. *Language* 81(2): 345–381.

Lasersohn, Peter. 1999. Pragmatic halos. *Language* 75(3): 522–551.

Lewis, David. 1969. *Convention*. Cambridge, MA: Harvard University Press. Reprinted 2002 by Blackwell.

Potts, Christopher. 2008. Review article: Hagit Borer's *Structuring Sense, Volumes I and II*. *Language* 84(2).

Schelling, Thomas C. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Syrett, Kristen, Evan Bradley, Christopher Kennedy & Jeffrey Lidz. 2005. Shifting standards: Children's understanding of gradable adjectives. URL http://semantics.uchicago.edu/kennedy/docs/sbkl-galana05.html. Paper presented at the 2005 Annual Meeting of the Linguistic Society of America, San Francisco, CA.

Taylor, Peter D. & Leo B. Jonker. 1978. Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences* 40(1): 145–156.