The Textual Shadow of Language: Speech Acts as Effects, and the Realizability Gap of Text-Only Models

PIATRA . INSTITUTE

July 2026

Abstract

A simulation of water does not wet, and a model trained on text does not, by that training, perform the speech acts whose records it reproduces. This paper turns the intuition into a theorem with a worked example. Full language is an effectful process. A speech act, in the sense of Austin and Searle, carries a context and a world-state, the social, epistemic, and normative situation, to an updated world-state together with an utterance, so the complete object is a state transformer $s \mapsto (s', u)$ and the utterance u is only its emitted value. Text records u and drops the transition $s \mapsto s'$. The textual shadow of a linguistic process is defined here as the observational collapse that keeps only textual outcomes, restricts to textual probes, and quotients the world-states that no textual probe can separate; it is the biextensional collapse of an associated Chu space and, for a stochastic token stream, the causal-state machine of computational mechanics. A model fit to text recovers that shadow and nothing finer. The consequence is an underdetermination result: distinct effectful processes can induce one and the same textual shadow, so no functional of the text distribution recovers which process cast it, and reproducing strings is therefore not performing the acts. A deterministic computation exhibits the gap. Two speech-act processes over a shared two-token alphabet, one binding every promise it utters and one binding none, are built so that their token-sequence distributions agree to machine precision while their effects on world-state diverge. The shared shadow is a two-state predictive machine of statistical complexity about 0.985 bits and entropy rate about 0.920 bits per token; after 100 tokens the two processes hold an expected 42.86 outstanding obligations against 0, a whole interval of world-states consistent with a single text; a third process makes each promise binding on a hidden fair coin, leaving 1 bit of illocutionary information per promise that no text-only learner can recover, and merging two distinct pragmatic states onto a single predictive state. The strongest opposing view, that the shadow preserves far more than critics grant because inferential and pragmatic structure leaves textual traces, is granted in full, and the residue it does not reach is located exactly: the world-transition the utterance was the act of producing. The boundary is a claim about recovery rather than a verdict on understanding. What closes the gap is grounding, interaction, and a handler in the world that gives an utterance its force.

A simulation of language does not speak

A simulation of water does not wet. The slogan is built to humble a program that produces something, and it half-succeeds. Wetness is a property of water against a surface, and a simulation of the molecular dynamics computes the property without instantiating it. Language resists the analogy at first, because language is not a substance and tolerates many substrates. The same sentence rides on vocal folds, ink, a shaped hand in sign, a pattern of current, or a run of tokens from a model, and it stays language when silicon rather than a larynx emits it. A system that produces well-formed, apposite, context-sensitive sentences is producing text, and the text is real.

The question here is narrower and has an answer. Grant that the output is language in the thin sense of being text of the right shape. What does a system trained on text alone thereby recover of language in the thick sense, the sense in which an official saying “I now pronounce you married”

changes who is married to whom? Two replies come quickly and both overshoot. One holds that the model copies surface strings and recovers nothing; the octopus of Bender and Koller (2020), eavesdropping on a telegraph cable it cannot connect to the world, learns only which words tend to trail which. The other holds that producing appropriate language simply is linguistic competence, so a fluent model is a speaker. The first reply understates what the statistics of text contain. The second forgets what text omits.

The position defended here sits between the two and is exact about the seam. A speech act changes the world, and text keeps the words while dropping the change. Call the record of the words the *textual shadow* of the act. A model fit to text learns the shadow, the minimal machine that predicts the next word from the words so far, and the shadow is a coarse projection of the act that cast it, having discarded the effect and merged pragmatic situations that the words alone cannot tell apart. Three claims make this precise and give it teeth. The textual shadow admits a clean definition as an observational collapse. A model fit to text provably recovers that collapse and no more, in the specific form of the predictive states of the text stream. And distinct linguistic processes can share a shadow, so text-only fitting cannot in principle recover which process produced the text.

One example runs through the paper. A speaker says, "I promise to repay you." The four words are fixed. Behind them the situation may be a sincere promise that puts the speaker under an obligation, an insincere one that feigns the obligation, a line quoted from a play, an utterance by someone with no standing to bind anyone, a continuation generated by a model, or an action taken by a system an institution has authorized to commit it. The text is identical across all six. What differs is what each does to the world once uttered, and that difference is the subject of this paper.

The utterance is the return value of an act

Austin's lectures took performative utterances as their starting case: to say "I promise," "I bequeath," "I name this ship," under the right conditions, is not to report that one promises or bequeaths or names, but to do it (Austin, 1962). The utterance is an instrument, and the act it performs is a change in a shared situation, an obligation incurred, a title transferred, a name conferred. Austin's felicity conditions govern whether the change takes: the wrong speaker, the wrong setting, the wrong uptake, and the words come out while the deed misfires. Searle regimented the picture into illocutionary acts constituted by rules, where an assertion places its speaker under a commitment to truth, a promise creates an obligation, and a declaration alters an institutional fact by being performed (Searle, 1969). A sincerity condition separates the sincere promise from the insincere one; the two share every word and differ in the state they leave behind.

The structure Austin and Searle describe is the structure of a computational effect. An effect is what an operation does besides return a value: it reads or writes a store, raises an exception, consults an environment, changes a state that outlives the call. Plotkin and Power model such operations as the generators of an algebraic theory acting on a state-carrying monad, so that a computation is a term over those operations and its meaning is the state transformation it induces (Plotkin and Power, 2003). A speech act fits the mould. Let S be the world-state, meaning the social, epistemic, institutional, and normative situation, and let U be the space of utterances. A speech act is a map

$$\text{say} : S \longrightarrow T(S \times U),$$

a state transformer that returns an utterance and, as its effect, updates the state, with T recording whatever nondeterminism or partiality the setting carries. Uttering "I promise to repay you" sends a state s to a pair: the string, and a state s' in which an obligation now stands. A judge's "guilty" returns a word and installs a legal status. A command from an authorized officer returns a word and changes what is permitted.

Text is one projection of this object. Compose with the map that keeps the utterance and forgets the state,

$$q : T(S \times U) \longrightarrow U,$$

and what remains is a context-to-string function, the thing a language model is trained to imitate. The projection is lossy by construction. It retains the return value and deletes the effect. A model can reproduce u with high fidelity while instantiating nothing of $s \mapsto s'$, so it can say "I promise" while no obligation comes to hold. A trace is a recorded string. An act is a state transformer. A process is a system of such transformers unfolding over a history. The three are different objects, and text is the shadow of the third under the projection that keeps only the first.

The textual shadow as an observational collapse

Two independent losses separate the act from its record, and it helps to take them one at a time before naming the object they produce. The first is a restriction of probes. A full linguistic situation admits many kinds of probe: one can read what was said, and also check what the speaker then owed, what became permitted, what the hearer came to know, what an institution now recognized. Text keeps only the probes that read emitted strings and drops the rest. The second is a coarsening of outcomes. Even under a textual probe, the full outcome of an interaction is a tuple of an utterance and a bundle of world-transitions; the projection q keeps the utterance and discards the transitions. Restriction throws away questions. Coarsening throws away parts of each answer.

After both losses, distinct world-states can become indistinguishable. Represent a linguistic system as an evaluation $r : S \times C \to K$ of world-states S against contexts C returning outcomes $K = U \times W$, where U is the utterance and W the world-transition. Fix the textual contexts $C_{\text{text}} \subseteq C$ and the projection $q : K \to U$, and write $r_{\text{text}}(s, c) = q\big(r(s, c)\big)$ for the textual evaluation.

Definition 1 (Textual shadow)**.** *Two world-states are textually equivalent, $s \sim_{text} s'$, when $r_{text}(s, c) = r_{text}(s', c)$ for every textual context $c \in C_{text}$. The* textual shadow *of the system is the quotient* $\text{Sh} = (S/\!\sim_{text},\ \bar{r}_{text},\ C_{text})$ *carrying the induced evaluation, with any two textually equivalent contexts likewise identified.*

The quotient is a standard construction wearing a linguistic interpretation. Presented as a table of states against textual probes with string-valued entries, it is a Chu space, and the identification of

duplicate rows and duplicate columns is the biextensional collapse that removes observationally redundant states and probes (Barr, 1996). Presented as a system whose behaviour is what observation can elicit, it is the quotient by observational equivalence, where two states are identified when no experiment separates them, and the quotient is the minimal realization of the observable behaviour (Rutten, 2000). Either way the shadow keeps a distinction if and only if some textual probe registers it, and discards every distinction that only a non-textual probe, or the withheld world-transition, could have shown. A sincere promise and a quoted one, a binding declaration and a powerless one, fall together in the collapse when the words they emit coincide.

What the shadow keeps and what it drops

The projection would be uninteresting if it kept little, and the strongest case against the present thesis is that it keeps almost everything. Much of what looks like hidden pragmatic structure leaves textual traces, and a model that reads enough text recovers a great deal of it. Grice showed that what a speaker implies beyond what they say is computed from the said together with the assumption that they are cooperating, so implicature is reconstructible from textual context and the surrounding discourse rather than from anything outside it (Grice, 1975). Wittgenstein located meaning in use within a language-game, and use is largely public and largely on the record, so a corpus of use carries much of what there is to know (Wittgenstein, 1953). Brandom built meaning from the deontic scorekeeping of a conversation, the commitments and entitlements that speakers attribute and undertake as they reason with one another, and that scorekeeping is articulated in what is said and asked and challenged (Brandom, 1994). An inferentialist has real grounds to expect that a model trained on the giving and asking for reasons will internalize much of the inferential articulation of concepts, because that articulation is worn on the surface of discourse.

Grant the case at full strength. The distributional structure of text is rich, and a system that models it well captures implicature, inferential role, genre, register, and the fine conditional expectations that let one word predict the next. Harris, stating the hypothesis such models rest on, described a language as fully specifiable by the distribution of its parts relative to one another, and was candid about the price of the specification, that it proceeds without intrusion of other features such as history or meaning (Harris, 1954). That candor marks the seam. The distributional structure is everything the shadow keeps, and it is a lot. The world-transition W is what it drops, and no amount of distributional richness reconstitutes W, because q deleted it before the statistics were ever taken. Whether a given "I promise" left an obligation standing is a fact about $s \mapsto s'$, and two histories that agree on every word while disagreeing on that fact are, to the shadow, the same history. The steelman is right. The shadow is large, and it is silent on the one component the projection was defined to remove.

The model learns the minimal predictive machine of the shadow

For a stochastic token stream the shadow has a canonical form, and it is the object a language model approximates. Computational mechanics defines the *causal states* of a process as the classes of pasts that induce the same conditional distribution over futures: two histories are one causal

state when they license the same predictions. The machine that moves between causal states as new symbols arrive, the epsilon-machine, is the unique minimal model that is maximally predictive of the process, and the entropy of its stationary state distribution, the statistical complexity, measures how much of the past the process must store to predict its future as well as the future can be predicted (Shalizi and Crutchfield, 2001). The causal states are the coarsest summary of history that loses nothing about what comes next. They are a sufficient statistic for the future.

This is the textual shadow of a token stream, made concrete. A model trained to predict the next token from the preceding ones is fit to the conditional-future structure of the text and to nothing else, so its sufficient internal state is, at best, an approximate coordinate for the causal states of the token process. Its representation is predictive of text to the degree that it recovers those states, and predictive sufficiency for text is a weaker property than sufficiency for the act. A coordinate that predicts the next word perfectly can be blind to whether the last promise bound anyone, because the binding never entered the token statistics it was fit to. The model recovers the causal-state machine of the shadow. The world-transition sits outside that machine, in the component the projection removed.

The construction of the next section makes the machine explicit. Its token process is an order-1 Markov chain on a two-symbol alphabet, and reconstructing its causal states from the token statistics alone yields a two-state predictive machine, one state per most-recent symbol, with statistical complexity about 0.985 bits and entropy rate about 0.920 bits per token. Those two numbers exhaust what the shadow has to say about the stream. Everything the effectful process does to the world is invisible to them.

Distinct processes, one shadow

The boundary can be stated as a proposition, kept deliberately modest. It forbids nothing about what a model might do when coupled to the world. It states only that one kind of information, the world-transition of a linguistic act, is not a function of the text the act emits, so a learner with access to the text distribution alone cannot recover it.

Proposition 1 (Textual underdetermination)**.** *There exist distinct linguistic processes that agree in their textual evaluation and disagree in their world-transition. Any two such processes have the same textual shadow. Consequently every functional of the text distribution takes equal values on them, and no learner fit to text alone can recover the world-transition.*

Proof. Let $\mathfrak{L}_1$ and $\mathfrak{L}_2$ share a state space and a textual evaluation r_{text} while carrying world-transitions $W_1 \neq W_2$; the construction below exhibits such a pair, in which the projection q discards W and the two processes assign identical probabilities to every token sequence. Because $\sim_{\text{text}}$ and $\bar{r}_{\text{text}}$ depend on r_{text} alone, the shadows $\text{Sh}(\mathfrak{L}_1)$ and $\text{Sh}(\mathfrak{L}_2)$ coincide. A learner with access only to text computes a functional F that factors through the shadow, so $F(\mathfrak{L}_1) = F(\text{Sh}(\mathfrak{L}_1)) = F(\text{Sh}(\mathfrak{L}_2)) = F(\mathfrak{L}_2)$, and F cannot separate W_1 from W_2. The world-transition is not among the quantities text determines. □

A deterministic computation makes the pair concrete. It is illustrative, constructed to instantiate the definitions above rather than sampled from natural language, and every number it reports is an exact property of the constructed processes, with nothing drawn at random. The token alphabet has two symbols, an ordinary assertion a and the promissive p reading "I promise to repay you." A single emission process governs the tokens, an order-1 Markov chain whose stationary distribution places about 0.571 of its mass on a and about 0.429 on p. Above this one emission process sit three speech-act processes that differ only in what a promise does to the world. The binding process puts the speaker under an obligation on every p. The detached process, the parrot with no standing, creates no obligation on any p. The hidden-coin process makes each promise binding or empty on a fair coin that never surfaces in the tokens. Because the effect variable never feeds back into emission, all three assign identical probabilities to every token sequence, and the computation confirms the agreement across all sequences up to length 8, where the largest probability difference falls below 10^{-15}.

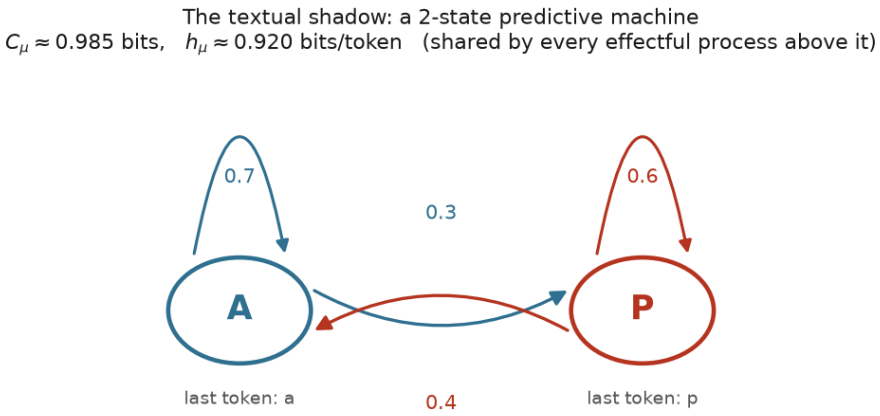

Figure 1: The textual shadow of the constructed stream: the two-state predictive machine reconstructed from the token statistics alone, with statistical complexity about 0.985 bits and entropy rate about 0.920 bits per token. The binding, detached, and hidden-coin processes all project onto this one machine.

The three processes therefore share one textual shadow, the two-state machine of the previous section, and a learner fit to the text cannot tell them apart. Their effects on the world do not agree at all. Counting an incurred obligation as one unit of world-state, the binding process accumulates an expected 42.86 outstanding obligations over 100 tokens, the detached process accumulates 0, and the hidden-coin process accumulates an expected 21.43. Sweeping the sincerity rate from 0 to 1 traces the whole interval from 0 to 42.86: every value in it is the world-state of some effectful process consistent with the very same distribution over texts. The text fixes a point in the space of shadows and leaves an entire interval of worlds standing behind it. The record underdetermines the world.

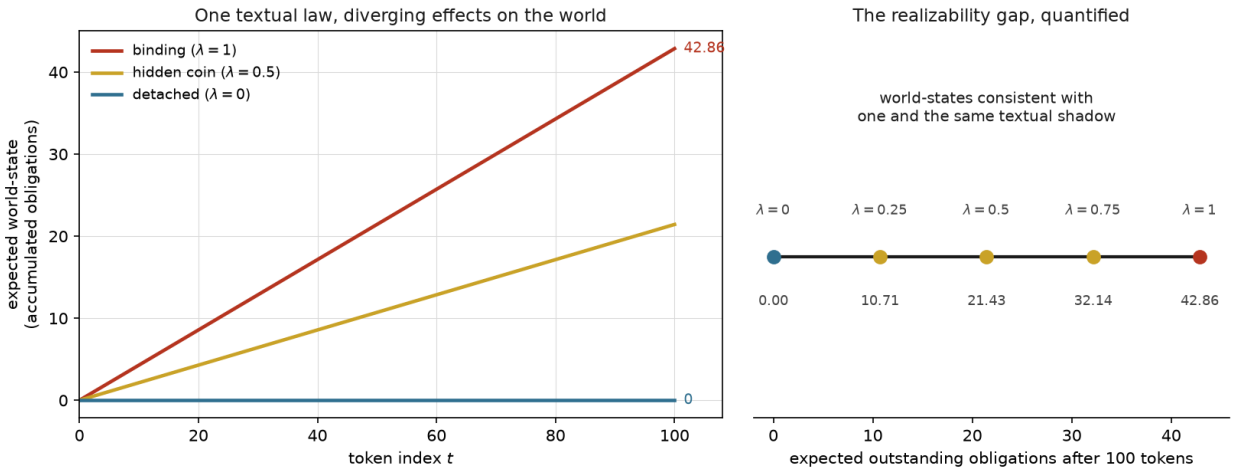

Figure 2: One textual law, diverging effects. Left: the expected accumulated world-state under three processes that share every token-sequence probability; from an identical text the worlds separate. Right: the interval of world-states consistent with that one text, spanned by the sincerity rate, from 0 under the detached process to 42.86 under the binding one.

The hidden-coin process sharpens the point from underdetermination across processes to irrecoverability within one. Whether a given promise binds is a fair coin the tokens never reveal, so the text leaves 1 bit of illocutionary information undetermined at each promise, about 42.86 bits over 100 tokens that no text-only learner can supply. The same process shows the collapse directly. Augmenting its state with a single bit for whether an obligation currently stands gives three reachable pragmatic states, which the shadow's two causal states must merge: the predictive state reached after a promise is the image of two distinct pragmatic states, one owing and one not, and the shadow leaves 1 bit of obligation status undecided there. Two situations the world holds apart are, to the record, one.

Illocutionary force is supplied by a handler

If text does not carry the effect, something else must, and the algebraic account of effects names it. In that account an operation is inert syntax until a *handler* interprets it into a concrete model, and the same operation acquires different meanings under different handlers; a raised exception does nothing until a surrounding handler decides what catching it amounts to (Plotkin and Pretnar, 2013). Read speech this way, an utterance is a request to perform an illocutionary operation, declare(married) or promise(repay), and its force is whatever the ambient institution, standing, and context make of the request. There is no context-free function from strings to acts. There is an operation, and there is a handler,

$$\mathrm{Force}(u;\, C, r, I) \;=\; H_{C,r,I}(\mathrm{op}(u)),$$

that turns the operation $\mathrm{op}(u)$ requested by the utterance into an effect relative to a context C, a

role r, and an institution I. The declaration declare(married) handled by an authorized officiant returns a change of marital status, handled by a bystander returns nothing but air, handled by a theatre returns a fictional event inside the play, and handled by a model with no institutional coupling returns text and text alone.

This relocates the question "when does text perform an act" to a place where it has an answer. The utterance performs the act when a handler in the world binds it to an effect, and not otherwise. A language model, on its own, emits the free syntax of illocutionary operations with great fluency and supplies no handler, so its "I promise" is a well-formed request that nothing discharges. Coupling the model to a system that can act, an account it can debit, an interface it can call, a role an institution has granted it, installs a handler, and the same string then has a force it lacked before. The force was never in the string. It is in the handler, and the handler lives in the world, outside the text and outside the model that emits it.

What would close the gap

More text does not close the gap. Scaling the corpus refines the shadow, sharpens the predictive machine, and lowers the entropy rate the model has yet to reach, and it does none of this to the world-transition, which was projected out before the first token was counted and stays absent at every scale. The missing datum was never written into the text, so no quantity of text contains it. Three things supply it, and each reaches outside the token stream. Grounding connects symbols to non-symbolic contact with the world, which is Harnad's requirement that the meanings of a symbol system be anchored in sensorimotor capacity rather than left parasitic on the interpreters who read the symbols (Harnad, 1990). Interaction restores the probes the shadow dropped, letting a system be tested by what it does and not only by what it says, the resource the octopus lacked when it could hear the cable and never pull on the world at the far end (Bender and Koller, 2020). A handler binds emitted utterances to effects, so a promise the system makes can put it under an obligation an institution will enforce. Each addition returns information the projection removed, and none is a property a larger text corpus can acquire.

This is a claim about recovery, and it should be kept apart from the older claim it resembles. Searle's Chinese Room argues that symbol manipulation is not understanding, a thesis about the intrinsic intentionality of a system and about what it is like, or not like, to be one (Searle, 1980). The present result says nothing about understanding or experience. It says that the world-transition of a speech act is not a function of the act's textual output, so a learner given the output cannot recover the transition, whatever is or is not going on inside the learner. A system that did understand, in any sense one likes, would face the same barrier if it were shown only text, because the barrier is in the projection and not in the reader. The point survives every position in the philosophy of mind, which is why it is worth isolating from them. Bender and colleagues named a system that stitches text from distributional statistics without reference to meaning a stochastic parrot (Bender, Gebru, McMillan-Major, and Shmitchell, 2021). "Shadow" is more exact than "parrot," because a shadow is a definite projection of a definite body, it preserves real structure faithfully, and it is missing one dimension only: the one along the light.

What follows is practical. To deploy a language model as an agent that promises, certifies, orders, or declares is to install a handler that binds its utterances to consequences in the world. That installation is an institutional act performed by whoever wires the model to the account, the registry, the actuator, or the office, and it is where the force of the model's speech is decided and where responsibility for that force comes to rest. The text the model emits is the shadow. The handler is the body that casts it. Reading the shadow more closely will never show the body, and taking the one for the other is not a technical error about models but a category error about where in the world acts are performed.

References

Austin, J. L. (1962). *How to Do Things with Words*. Oxford: Clarendon Press.

Barr, M. (1996). The Chu construction. *Theory and Applications of Categories*, 2(2), 17–35.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). New York: Association for Computing Machinery.

Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5185–5198). Association for Computational Linguistics.

Brandom, R. B. (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics, Volume 3: Speech Acts* (pp. 41–58). New York: Academic Press.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1–3), 335–346.

Harris, Z. S. (1954). Distributional structure. *Word*, 10(2–3), 146–162.

Plotkin, G., & Power, J. (2003). Algebraic operations and generic effects. *Applied Categorical Structures*, 11(1), 69–94.

Plotkin, G., & Pretnar, M. (2013). Handling algebraic effects. *Logical Methods in Computer Science*, 9(4), Article 23.

Rutten, J. J. M. M. (2000). Universal coalgebra: A theory of systems. *Theoretical Computer Science*, 249(1), 3–80.

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–424.

Shalizi, C. R., & Crutchfield, J. P. (2001). Computational mechanics: Pattern and prediction, structure and simplicity. *Journal of Statistical Physics*, 104(3), 817–879.

Wittgenstein, L. (1953). *Philosophical Investigations* (G. E. M. Anscombe, Trans.). Oxford: Blackwell.