The Evolution of Denial

Luca Incurvati and Giorgio Sbardolini

To appear in *The British Journal for the Philosophy of Science*. Please cite published version.

Abstract

Negation is common to all human languages. What explains its universality? Our hypothesis is that the emergence of expressions for denial, such as the word 'not', is an adaptation to existing conditions in the social and informational environment: a specific linguistic form was co-opted to express denial, given a preference for information sharing, the limits of a finite lexicon, and localized social repercussions against synonymy. In support of our hypothesis, we present a costly signalling model of communication. The model formalizes ordinary aspects of Stalnakerian conversations, implements the conditions we isolated for the emergence of denial, and computes their long-term consequences through a widely employed evolutionary dynamics, whose results are calculated by computer simulations. The model shows that, under a reasonable configuration of parameter values, functional pressure derived from conversational constraints favours the emergence of denial by means of a dedicated expression, such as the word 'not'.

1 Denial is Adaptive

A linguistic form to express denial (what we typically do with negation) is common to all human languages, and absent from any other communication system of the natural world (Horn [1989]; De Swart [2010]). What explains the universality of negation in human languages? It is obvious that negation is very useful. However, not everything useful actually evolved: opposable thumbs are quite useful, but we are the only primates to have them. Their development appears to be related to the development of our larger brain, to the erect posture, which liberates the upper arms from the need to support movement, and therefore to our life on the ground, unlike that of other primates. An evolutionary argument from usefulness (or rather, fitness) can explain the universality of negation only in the presence of plausible additional assumptions.

Our hypothesis is that the emergence of expressions for denial, for example through the word 'not', is an adaptation to existing conditions in the social and informational environment. In support of this hypothesis, we formulate a set of such conditions. First, denial has to be advantageous in relatively simple scenarios, which plausibly have taken place frequently enough in the course of our history to sustain universal development. Second, the innovation must be initially relatively small, otherwise there would be a sudden jump in nature. Finally, it could not have been too costly to maintain it, lest its postulation be unjustified. We specify these conditions in a game-theoretic model of cultural evolution, following a standard methodology for evolutionary studies in linguistics (Steels [2011]): we reconstruct ordinary speech situations as games, and draw from evolutionary game theory to study the emergence of behavioural regularities through time (Weibull [1995]; Hofbauer and Sigmund [1998]).[1]

Given the specified conditions, the model predicts the emergence of denial under a standard evolutionary dynamics. The resulting account is essentially Darwinian in its logic: large-scale results observable in the grammar of human languages are extrapolated from short-term effects of everyday conversations (Croft [2000]). In the context of more general adaptationist arguments for the evolution of language (Pinker and Bloom [1990]; Pinker and Jackendoff [2005]; Hauser *et al.* [2014]), we focus on the local question of the evolution of negation. A specific linguistic form was co-opted to express denial, given a preference for information-sharing, the limits of a finite lexicon, and localized social repercussions against synonymy. Functional necessity (in discourse) leads, and the evolution of form follows.

We begin in Section 2 by discussing properties that distinguish assertion and denial, and conditions for the emergence of the latter in a signalling game. In Section 3 we introduce the rejection game: a signalling game in which agents accept or reject information (Incurvati and Sbardolini, unpublished). In Section 4, we present our model of the evolution of denial, and give equilibrium results under the evolutionary dynamics we employ: the Wright-Fisher process with selection (Ewens [1979]), of which we give a cultural interpretation. In Section 5 we review our model,

[1] Unlike biological evolution, in cultural evolution replication is imitative and not sexual. There are various other differences between cultural and biological contexts, which may be reflected in different modelling assumptions. For a discussion of cultural evolution, especially in connection with the evolution of language, see (Boyd and Richerson [1985]; Croft [2000]; Blythe and Croft [2012]).

compare it with related work, and indicate some possible developments.

2 Signalling to Deny

A simple model of linguistic conventions is the signalling game (Lewis [1969]; Nowak and Krakauer [1999]; Skyrms [2010]; Huttegger and Zollman [2011]): a strategic interaction between at least two players, sender S and receiver L, which we can think of as Speaker and Listener. In the simplest non-trivial setting, S has two signals, m_1 and m_2, and is in one of two information states (or 'world states'), s_1 and s_2. L doesn't have S's information, but performs one of two actions, r_1 and r_2, on the basis of the signal received. Payoffs depend on whether L's action is appropriate relative to S's information. A strategy is a complete specification of a player's actions: pure strategies are, for S, functions from states I to signals M, and for L functions from M to actions R. More generally, for Δ a probability distribution, mixed strategies are functions $I \mapsto \Delta M$ and $M \mapsto \Delta R$, for S and L respectively.

The players' preferences about the outcomes of their interactions are represented by payoffs, that is, numerical values. The description of the game is completed by a specification of how payoffs are rewarded to an agent who follows a given strategy in a given state. This is done by a utility function. A (state-dependent and discrete) utility function u_i for player i (one of S and L) is a function from a state s and an action r to a payoff: either 0 or a, for some $a > 0$. Let σ_S and σ_L be S's and L's strategy, respectively. The expected utility of their combination for player i is the utility u_i of performing action r in state s weighted by the probability of s, the probability that S sends m if s is observed, and the probability that L performs r if m is sent. In symbols:

$$eu_i(\sigma_S, \sigma_L) = \sum_{s \in I} \sum_{m \in M} \sum_{r \in R} Pr(s) \cdot \sigma_S(s) \cdot \sigma_L(m) \cdot u_i(s, r)$$

Some combinations of strategies may be better than others. Let σ_i be i's strategy, and σ_{-i} an ordered list of everyone's strategies except for i's. Then the pair $\langle \sigma_i, \sigma_{-i} \rangle$ is a (weak) Nash equilibrium if, and only if, for all players i and all strategies $\sigma'_i \neq \sigma_i$, $eu_i(\sigma_i, \sigma_{-i}) \geq eu_i(\sigma'_i, \sigma_{-i})$. In words, a Nash equilibrium is a combination of strategies such that no player would prefer to play another strategy, other things being equal (Osborne and Rubinstein [1994]).

A Nash equilibrium in a signalling game can be understood as solving a communication problem (Lewis [1969]): the problem of using signals to transfer information. Typically, reliable information transfer is possible if the players are coordinators—if their preferences are aligned over the same outcomes—though some degree of communication is marginally possible with imperfect alignment of preferences (Crawford and Sobel [1982]; Martinez and Godfrey-Smith [2016]). It is common to assume for the study of communication that L's action consists in forming a belief. Coordination obtains just in case L's final belief coincides with S's initial information.

$$\text{for all players } i, u_i(s, r) = \begin{cases} a & \text{if } s = r \\ 0 & \text{otherwise} \end{cases}$$

$$
\begin{tikzcd}
s_1 \arrow[r] & m_1 \arrow[r] & r_1 \\
s_2 \arrow[r] & m_2 \arrow[r] & r_2
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
s_1 \arrow[dr] & m_1 \arrow[dr] & r_1 \\
s_2 \arrow[ur] & m_2 \arrow[ur] & r_2
\end{tikzcd}
$$

Figure 1: Nash equilibria in pure strategies in a 2x2x2 signalling game

Figure 1 represents the only two Nash equilibria in pure strategies of the signalling game with two states, two signals, and two actions, played by coordinators.

Lewis ([1969]) argued that, if either of these two outcomes obtains, S can be described as communicating her private information to L by symbolic means. If the players settle on the option on the left in Figure 1, m_1 is used to signal s_1 (= r_1) and m_2 to signal s_2 (= r_2). If the players settle on the option on the right, the opposite is the case. There is no ambiguity, and no overlap, in the use and interpretation of signals: these equilibria are called separating.[2] According to Lewis, the possibility of multiple equally good solutions to the agents' communication problem accounts for the conventionality of human languages.

There are many other possible strategies for S and L besides those in Figure 1, so it is not obvious how the players should converge on them. The problem of equilibrium selection has been addressed in models of cultural and biological evolution, which show that small incremental advantages can accumulate through repeated interactions over time (Huttegger [2007]; Skyrms [2010]). From an evolutionary perspective, only relatively modest cognitive assumptions are made about the players: they must be capable of receiving feedback from experience, and have limited retention of advantages accrued over time (Barrett and Zollman [2009]; Spike *et al.* [2017]).

The resulting picture is very attractive: assuming a deflationary conception of meaning as positive signal-states correlation (somewhat along the lines of Grice's ([1989]) natural meaning), rather unsophisticated cognitive abilities suffice to establish a communication system in which signals with no pre-existing meanings are reliably used to sort different information states. The next challenge is to devise a signalling system that models negation.

2.1 Denial and costs

Intuitively, sending a signal corresponds to asserting a message. Drawing on a philosophical tradition that includes Price ([1990]), Smiley ([1996]), Rumfitt ([2000]), and Incurvati and Schlöder ([2017]) among others, we consider negation as a device that may be attached to a signal m, such that sending the whole, $-m$ (the signal with negation), corresponds to denying a message. From this perspective, we introduce some conditions that, over evolutionary time, establish a difference between signalling to assert and signalling to deny. We implement these conditions in a costly signalling model.

In biology and economics, costly signalling is a common account of what keeps communication

[2] In this case, all Nash equilibria are separating, but this is an artifact of simplicity. In general, the situation is more complex. See (Pawlowitsch [2008]; Huttegger *et al.* [2010]).

honest despite conflicting interests among interlocutors. If there is an incentive to deceive, signals that come with a cost may be too much to send for a dishonest signaller, for the gains of deception may not compensate for the costs. There are two types of costs: production costs and social costs. We will employ both.

Production costs are often associated with Zahavi's ([1975]) handicap principle. Elaborate secondary sexual characters may be a hindrance to male survival, but, according to Zahavi, they are nevertheless selected in their function as signals of superior fitness targeting the female. The development and maintenance of the peacock's long tail, for example, takes a toll on the male carrying it, but since weaker males can't bear the otherwise useless display, sexual selection benefits those who can. The handicap principle remains controversial (Penn and Számadó [2019]), but the idea of 'honesty through waste' has been generally accepted in economics (Veblen [1899]; Spence [1973]). As Lachmann *et al.* ([2001]) point out, the explanatory force of the handicap principle is attached to signalling that influences the behaviour of a signal receiver, whose choices have effects that only become visible in the long run. The peahen has no immediate feedback for her choice of a mate, which factors in as offspring survival only over a generation. These conditions may lead to achieving honest communication by showing off wastefully.

Dishonest signalling may also have more immediate consequences, for example if the signal receiver can readily verify the appropriateness of the signal, in which case the signal sender should be prepared to face the consequences of unveiled deception. Such negative consequences may be factored in as social costs (van Rooij [2003]). Maynard Smith and Harper ([1988]) apply a social cost model to account for dominance behaviour in the common sparrow. In this case, the costs of producing a signal of aggressiveness are negligible, but the bird threatening to attack should be prepared in case the signal receiver (in this case, another bird) ignores the threat and fights: weaker animals may pay significant costs for dominance displays that they can't support in actual confrontation. So honest signalling does not require waste if the costs of deception are socially enforced immediately after the signal is sent (Maynard Smith [1991]; Lachmann *et al.* [2001]; Zollman *et al.* [2013]).

If the distinction between signalling $-m$ and signalling m must capture the difference between assertion and denial, at least the following properties should obtain: markedness, incompatibility and compositionality. These are properties an operator ought to have in order to be recognized as negation (even though, as we shall see, natural language negation is inevitably more complex). We now explain these properties and the corresponding conditions in the game that, on our hypothesis, facilitate their emergence. The first two conditions take the form of differential costs (that is, a combination of production and social costs), the third of a structural constraint on learnability.[3]

The first property, markedness, is supported by overwhelming linguistic and psychological evidence. Negation, or more generally denial, is marked.[4]

[3]For a definition of negation informed by a discussion of linguistic typology, see (Bond [2012]). Bond motivates a semantic characterization of negation in terms of binary contrast and exclusion which is closely related to our notion of incompatibility (see below).

[4]For general introductions to the notion of markedness in linguistics, see (Kean [1992]; Bybee [2011]). The concept of markedness encompasses phonology, syntax, and pragmatics. For its application in relation to negation in particular,

Markedness: Expressions used to deny are marked. Denial takes relatively more cognitive effort than assertion.

Use of an unmarked signal, for instance an utterance of (1a), isn't *ipso facto* a denial (that Fred is sad), lest we forfeit the difference between assertion and denial. Use of a marked signal, (1b), rather than an unmarked (near-)synonym such as (1a), tends to indicate that the unmarked alternative is less relevant for interpretation. Consequently, (1b) is usually taken to communicate that the Fred is not exactly happy (Grice [1975], [1989]; Sperber and Wilson [1995]; Levinson [2000]).

(1) a. Fred is happy.
b. Fred is not sad.

To account for markedness, we model denial as a production cost, which can be interpreted as the extra psycholinguistic and articulatory effort required to produce a denial. For many natural languages, it can be seen as the cost associated with the morphological and phonological realization of negation. We assume that for every 'basic' signal m, the speaker has the option of sending its denial $-m$, that is the signal with a small cost attached. The presence of a small production cost limits the use of costly signals to less frequent (less 'typical') situations and this is an aspect of the use of marked expressions in natural language that we recover.

Morphological modification of a word, in general, may express a variety of modulations on referential meaning (for example tense, aspect, number, gender, case). The next property of denial requires that the relation between a signal m and its costly counterpart $-m$ is incompatibility.

Incompatibility: Denial conveys incompatibility. To say that 'Fred is not in the kitchen' is to rule out that Fred is in the kitchen.

Incompatibility is a widespread assumption about negation (Price [1990]; Berto and Restall [2019]; Bond [2012]). We will assume that consistency of signalling behaviour (you cannot say 'It's raining' and 'It's not raining' in the same breath) is socially enforced: social costs are inflicted on the signaller who sends m and $-m$ in the same state. We call this assumption polarity.

Polarity is in effect an anti-synonymy condition. Synonymy, or the use of different signals in the same state, is a widespread feature of natural language. Recent work has emphasized how many properties of language, including (some degree of) synonymy and ambiguity, are to be expected on the general assumption that a functionally organized communication system balances informativity and simplicity (Zipf [1949]; Manin [2008]; Piantadosi *et al.* [2012]). However, polarity does not apply across the board: speakers are not penalized for sending any two m and m' in the same state, only for sending m and its costly counterpart $-m$ in the same state. In other words, polarity is a specific assumption about the relationship between m and $-m$ (or, m' and $-m'$).

Besides incompatibility, denial has further logical properties. As we discuss in Section 5, some assumptions can be added to our account so that the denial of a sentence A is equivalent to the

see (Horn [1989]; De Swart [2010]).

assertion of its (classical) negation $\neg A$. Thus denial, as characterized in the present work, can be seen as the basis for the operator $\neg$ of classical logic (Smiley [1996]).

As far as Lewisian signalling games go, signals could be mere idioms. The challenge is to extend signalling theory to logically complex signals: negation attaches to a sentence to form another sentence whose meaning is determined compositionally.[5]

Compositionality: Expressions used to deny are semantically compositional. The meaning of 'Fred is not in the kitchen' is a function of the meaning of 'Fred is in the kitchen'.

There are various proposals about compositionality in signalling games (Barrett [2009]; Steinert-Threlkeld [2016]; Franke [2016]; LaCroix [forthcoming]). A key idea is that one advantage of syntactic complexity lies in greater expressive power, which can be achieved even on a small lexicon: if the information states to be distinguished outnumber the available atomic signals, it is only by combining the signals that agents may transfer information while still avoiding ambiguity (Batali [1998]; Kirby *et al.* [2015]). We share this perspective. Our implementation of compositional signalling follows, in its outline, the account of Franke ([2016]). We will introduce a relation ρ on the information states, and assume that rewards for sending m in a state s 'spill over' to sending $-m$ in a ρ-related state s'. In equilibrium, the resulting signalling system has the following property: the meaning of $-m$ is a function of the meaning of m.

This form of compositionality is fairly basic, in that it does not require symbol concatenation. In the end, we would like a more straightforwardly syntactical account of compositionality. We discuss the issue further in Section 5.

2.2 Denial and ordinary negation

The assumptions laid down so far fall short of 'ordinary' negation, which has many properties besides the ones above. For example, ordinary negation has rich scoping possibilities, it can be iterated, and it can be embedded under other operators. We will not account for these additional properties. Indeed, the signalling model does not show that the expression used for denial is recursive.

What we do describe is a form of denial that is present in nature, as evidence from Italian and Spanish shows. Moreover, it is plausible that something like this non-recursive form of denial occupies an intermediate level of development between signalling with no displacement (no way to signal about something other than what's present), and languages with a fully compositional, recursive, truth-functional negation operator. Small steps fill the gap between animal communication systems (such as primate alarm calls, or bird and cetacean songs) and human languages (Seyfarth *et al.* [1980]; Haldane [1992]; Zuberbühler [2002]; Hollén and Radford [2009]). Our model describes one of these intermediate stages.

[5] Various properties often go under the name 'compositionality'. See (Szabó [2000]; Pagin and Westerståhl [2010]) for discussion.

There are, of course, no direct records of the languages spoken by the first hominids. However, there may be 'fossils' of language evolution scattered through modern languages. Since recursion is a later development, perhaps exclusive to humans, Jackendoff suggests that, for example, the English response particles 'Yes!' and 'No!', which can't be fed to a recursive syntax, can be regarded as fossils in this sense: they are more primitive than other expressions with regular combinatorial properties (Jackendoff [1999]).

Jackendoff's suggestion is plausibly to be understood within a historical context: the existence of fossils in English does not imply that all languages underwent the same history. Since languages develop (somewhat) autonomously, observations about negation's origins must be relative to a typology of negation across languages (Dahl [2010]; Bond [2012]).[6] More generally, however, negative markers from earlier stages in the historical development of modern negation are familiar findings in reconstructions of Jespersen's cycle: a cross-linguistic process whereby the increase in frequency of an emphatic form leads to its standardization and to the recruitment of new forms for emphatic purposes (Jespersen [1917]; Dahl [1979]; Payne [1985]; van der Auwera [2009]).

In this regard, a particularly relevant piece of evidence comes from some varieties of Italian (Zanuttini [1997]; Frana and Rawlins [2019]) and Spanish (Negroni [2017]), in which we find a negative particle 'mica'/'minga' (Italian/Spanish, respectively) that (i) expresses denial, but (ii) cannot embed (hence doesn't fully exhibit the combinatorial properties of regular negation), and finally (iii) is an element of Jespersen's cycle (hence is evidence of language evolution).

The negative particle 'mica' exists in dialectal variants of Italian alongside regular sentential negation 'non'. According to Frana and Rawlins ([2019]), Italian 'mica' is a truth-value switcher, just like 'non', as well as a common-ground management operator, rejecting a presupposition of the interlocutor. In other words, 'Mica A' semantically asserts $\neg A$, and pragmatically rejects the interlocutor's prior belief that A. Example (2) shows that 'mica' expresses denial. Examples (3) and (4) illustrate the restricted scoping possibilities of 'mica'. Data for (2) and (3) are from (Frana and Rawlins [2019]).

(2) a. Non fa caldo.
'It's not hot'

b. Mica fa caldo.
Roughly: 'It's not hot' (what did you expect?!)

(3) a. Non devi guidare.
($\Box > \neg$) 'You must not drive';
($\neg > \Box$) 'You don't have to drive'

b. Mica devi guidare.
($\Box > \neg$) *'You must not drive';
($\neg > \Box$) Roughly: 'You don't have to drive' (what did you expect?!)

[6] For example, some languages, such as Welsh and Finnish, lack the equivalent of polar response particles 'Yes!' and 'No!' (Bloomfield and Hockett [1984], pp. 176–177). This does not refute Jackendoff's observation, as we understand it. Different languages have a different evolutionary history and may show different 'vestiges' of their past.

(4) a. Se non fa caldo, prenditi il cappotto.
'If it's not hot, take your coat'

b. *Se mica fa caldo, prenditi il cappotto.

Diachronically, 'mica' (from the Latin word for crumb) developed from a still existing negative intensifier ('non...mica': not...a bit) into a pre-verbal denial operator (Parry [2013]). Similar to 'mica', the marked expression described in our model (i) is a form of denial (ii) that lacks fully combinatorial properties. We describe a step in the gradual development of language, and the conditions under which an expression for denial with the properties listed above (markedness, incompatibility, compositionality) might enter Jespersen's cycle.

3 The Rejection Game

We have introduced the traditional signalling game (Lewis [1969]; Skyrms [2010]) and discussed some key properties of denial (markedness, incompatibility, compositionality) that we aim to account for. We also acknowledged that, while natural language negation is more complex, there are forms of denial which our model approximates and can be plausibly understood as early developments in the history of negation (such as Italian 'mica'). In this section, we introduce the particular game on which our model is based: the rejection game. In the next section, we add a system of costs to the rejection game to account for the difference between assertion and denial, and discuss under what conditions equilibria obtain.

The rejection game is designed to study Stalnakerian conversations (Stalnaker [1978]). In a Stalnakerian conversation, a speaker S asserts something, which a listener L accepts or rejects. Assertion is a proposal to update the common ground (information shared by S and L); the update goes into effect if the proposal is accepted, and is blocked if the proposal is rejected. We regard this type of interaction as fundamental to linguistic communication. A rejection game formalizes the decision-theoretic structure of this interaction: a multi-agent negotiation about what goes in the common ground (Incurvati and Sbardolini [unpublished]). Related games are studied in economics, often with somewhat different assumptions (Farrell and Rabin [1996]; Kamenica and Gentzkow [2011]).

Unlike Lewis/Skyrms signalling games, the rejection game is not played by coordinators: intuitively, there would be no point in rejection if what players do were inevitably in the interest of all. In a rejection game, S wins (maximizes payoffs) if and only if her signal is accepted, and L wins if and only if she accepts truths and rejects falsehoods. Since successful behaviour is reinforced, there is an incentive for S to deceive: if L successfully accepted m in the past, S might send it again, regardless of its truth, for L is more likely to accept it again.

Two kinds of game are compared in Figure 2. In both, player S has two actions, or signals to send, m_1 and m_2, and player L has two reactions, r_1 and r_2. As customary, payoffs for the column player are the first element in a pair of payoffs; and as above, payoffs are 0 and $a > 0$. The game on the left is a coordination game: a player wins if and only if the opponent does. The game on the right is the rejection game: on the assumption that m_1 is true and m_2 is false, r_1 and r_2 can

$$
\begin{array}{cc|cc}
 & & \multicolumn{2}{c}{S} \\
 & & m_1 & m_2 \\
\hline
\multirow{2}{*}{L} & r_1 & a,a & 0,0 \\
 & r_2 & 0,0 & a,a
\end{array}
\qquad\qquad
\begin{array}{cc|cc}
 & & \multicolumn{2}{c}{S} \\
 & & m_1 & m_2 \\
\hline
\multirow{2}{*}{L} & r_1 & a,a & a,0 \\
 & r_2 & 0,0 & 0,a
\end{array}
$$

Figure 2: Payoff structure of coordination game and hybrid game

be interpreted as acceptance and rejection, respectively. If S is truthful and sends m_1, L wins if and only if S does (they are coordinating). If S is not truthful and sends m_2, L wins if and only if S loses (they are in conflict). Rejection games are therefore hybrid: part coordination and part conflict.

Formally, a rejection game is a tuple consisting of a set of players $\{S, L\}$, a probability distribution Δ, a set I of information states, a set M of signals, a set $R = \{r_1, r_2\}$ of actions, and agent-relative utility functions $U_{S,L}$. The notions of (pure and mixed) strategy, expected utility, and of Nash equilibrium, are the same as above.

In a rejection game, S has an unconditional preference for L to take one of the two actions, say r_1, which is interpreted as acceptance. On the other hand, L prefers that S is truthful. Suppose that there are two initial information states, s_1 and s_2, and two signals, m_1 and m_2. Like Lewis ([1969]), we may assume that the meaning of the signals is arbitrary prior to the interaction, but that L prefers to receive m_1 if s_1, and m_2 if s_2, or else to receive m_2 if s_1, and m_1 if s_2. In other words, the listener prefers that the speaker avoids ambiguity. Basic reinforcement learning shows that speaker and listener often optimize their interaction: the listener always accepts, and the speaker always avoids ambiguity. If so, we observe the development of a conventional signalling system: the speaker sends the same signals for the same states, and different signals for different states (Incurvati and Sbardolini [unpublished]).

It is a common observation in economics that in hybrid games with cost-free signalling, such as the one in Figure 2 on the right, communication does not take place, precisely because there are no incentives for S to cooperate (Farrell and Rabin [1996]). After all, if L knows that S could be truthful or not, why should she ever accept what S says? Indeed, the game has a suboptimal equilibrium in which S signals randomly and L rejects everything.

As Sally ([2000; 2003]) points out, games of partial coordination and conflict can often be resolved if we assume that a player draws marginal gains from the other's win. This assumption is called sympathy, and may be measured by a parameter $0 \leq \lambda \leq 1$, which improves the speaker's payoff by a fraction of the listener's payoff. For $\lambda = 0$, we have a speaker who cares only that her signal is accepted; but for $\lambda \neq 0$, we have a speaker who partly benefits from the listener receiving unambiguous signals. Notice that $\lambda \neq 0$ does not imply that we are back into a coordination game, only that the speaker has a strict preference for the cooperative equilibrium. Thus, the use of sympathy balances the game towards cooperativity, and is particularly plausible when modelling conversations, which are typically cooperative interactions (Grice [1975; 1989]). In addition, from the perspective of our evolutionary argument, it is important to emphasize how recent research has found a central role for cooperativity as a source of the human ability to speak: an essentially interactive task (Clark [1996]; Tommasello [2003]; Scott-Phillips [2015]). The assumption that sympathy is a causal factor in the emergence of denial, insofar as it guides the speaker towards behaviour that favours the listener, is broadly in line with this perspective.

For the definition of utility functions for the rejection game, we first determine which signal is true in a state. Let s be a state, m a signal, r a reaction, and $a > 0$ a payoff. Let r_1 be S's preferred reaction (acceptance). Then $\mu(s)$ is the unique true message in state s just in case, intuitively, its acceptance in any other state would get the listener no benefit.

$$m = \mu(s) \text{ if, and only if, for all } s' \neq s, u_L(s', m, r_1) = 0$$

Thus, which signal is true in a state is initially arbitrary, and eventually determined by the interaction itself. Finally, we indicate a payoff b for successful rejection, with $0 < b \leq a$. We can now define utility functions: the speaker gets a positive reward a, plus anything in proportion to sympathy, just in case the listener accepts. The listener gets a if she accepts a true message, or if she rejects a false one.

$$\begin{aligned} u_S(s, m, r) &= \begin{cases} a + \lambda \cdot u_L(s, m, r) & \text{if } r = r_1 \\ 0 & \text{otherwise} \end{cases} \\ u_L(s, m, r) &= \begin{cases} a & \text{if } r = r_1, \text{ and } m = \mu(s) \\ a & \text{if } r = r_2, \text{ and } m \neq \mu(s) \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

If $\lambda = 0$, there is too much conflict for honest signalling to emerge, except by pure chance. But for $\lambda \neq 0$, it is easy to verify that the most optimal outcome for both players is one in which S uses different signals for different states, and L accepts them all. For if $\lambda \neq 0$, it is best for S that $u_L(s, m, r_1) > 0$ if L plays r_1, but this is only possible if for all $s' \neq s$, $u_L(s', m, r_1) = 0$, that is if for all other states, accepting the same signal would get L no benefit. Notice that acceptance and rejection are inversely correlated in the optimal equilibrium: a signal is accepted in a state just in case it is rejected in any other state, and *vice versa*.

For illustration, imagine that the two initial states are `predator` and `food`. The sentry signals to her companion whether the environment contains a predator or a food source, by sending one of two distinct calls m_1 or m_2. The companion prefers unambiguous calls, because she must act differently depending on what the environment is like: seek shelter if there's a predator, go grab the food if there's any. Thus, she gets a positive reward if, for example, (i) she accepts m_1 in `predator` and m_2 in `food`, and (ii) rejects m_2 in `predator` and m_1 in `food`, or *vice versa*. In this setting, the sentry gets a positive payoff simply for being listened to: if she deceives her companion and sends an anti-predator signal when in fact there is food around, her companion will run for safety, and the sentry can get a surplus of food. However, since she has some interest in her companion surviving (by reason of survival of the species), plausibly, $\lambda \neq 0$. This setting illustrates the two sides of a rejection game: cooperation, if the signal sender acts truthfully, and conflict, if

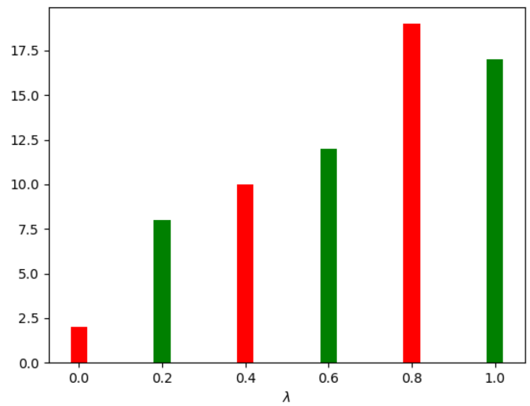

Figure 3: Results of rejection games for different values of λ.

she doesn't; but with sufficiently high sympathy, cooperative behaviour emerges (Incurvati and Sbardolini [unpublished]).[7]

Figure 3 shows the proportion of successful trials of the rejection game (with two states and two signals), over twenty-five repetitions, for different values of λ. A trial is a sequence of 500 rounds of the game, through the Wright-Fisher dynamics (see below), and is considered successful just in case, once terminated, the speaker signals unambiguously and the receiver accepts all, each with probability at least 0.8. (Trials are of arbitrary length, but 500 rounds are usually enough to find significant results.) As the figure shows, optimization tends to be proportional to the degree of cooperativity, and for $\lambda \geq 0.6$ the agents end up close enough to optimal signalling in the majority of trials.

A trial can be understood as a process of learning by trial and error, or of evolutionary development: in the course of this process the probability of playing a strategy at time t is determined by its probability at time $t-1$ and by how successful it is (Herrnstein [1970]). The more a strategy is successful (that is, the higher its expected utility) the more frequently it will be played at the next time step. Figure 4 plots the evolution of a rejection game during a sample successful trial, in this case with $\lambda = 0.5$, for a game with two states and two signals. Payoffs are set at $a = 3$. Probabilities are initialized by chance, and evolve non-trivially in different states according to how successful players are, given the Wright-Fisher dynamics. The meaning of m_1 and m_2 is initially arbitrary, but is built up through the interaction in the form of a robust signals-states correlation. Under Wright-Fisher, probabilities are never driven to 1 or 0, but keep oscillating due to inherent drift.

In an evolutionary setting, the correlation between expected utility and frequency is the standard

[7]The scenario is adapted from documented animal behaviour in the wild, in particular tufted capuchin monkeys (Wheeler [2009]). For a recent review of the growing body of work on animal anti-predator signalling, see (Hollén and Radford [2009]).

way to understand the fundamental notion of fitness, of which there are many mathematical models (Weibull [1995]). Some of these models are more readily interpreted as learning processes, others as evolutionary processes; either way, these models are to be considered fairly abstract descriptions of the phenomena. In our simulations, we employed a version of the Wright-Fisher model with selection, which is frequently applied to population genetics (Ewens [1979]).[8] As we explain in Appendix A, the Wright-Fisher process, without selection, is an elegant and general model of drift: random fluctuations in the distribution of a chosen value within a population of fixed size. In linguistics, drift is usually not interpreted as random variation in the copy of chromosomic information, but rather as occasional discrepancies in the imitation of linguistic behaviour, which is acquired by the infant from observation of adult speakers. In the long run, this 'cultural drift' may lead to macroscopic linguistic differences (Lassiter [2008]). On top of this, the model accounts for selection, by a coefficient that favours the survival of strategies with a higher relative fitness: that is, those types of behaviour that outperform alternatives in the rejection game. Thus, Wright-Fisher is particularly appropriate for evolutionary studies, as it allows to distinguish the unit of interaction, for which payoffs are defined (a linguistic agent), from the unit of replication, for which fitness is defined (a strategy type). Further discussion of the Wright-Fisher dynamics, as well as mathematical detail, may be found in Appendix A.

[8]For all the games discussed in this paper, we also ran simulations based on the replicator dynamics (Taylor and Jonker [1978]), another widely employed algorithm, and the results are similar to the ones reported. Wright-Fisher is a sophisticated choice in the context of evolutionary development, though it can be approximated by versions of the replicator dynamics that include mutation (Skyrms [2010], Chapter 5) and forgetting (Barrett and Zollman [2009]).

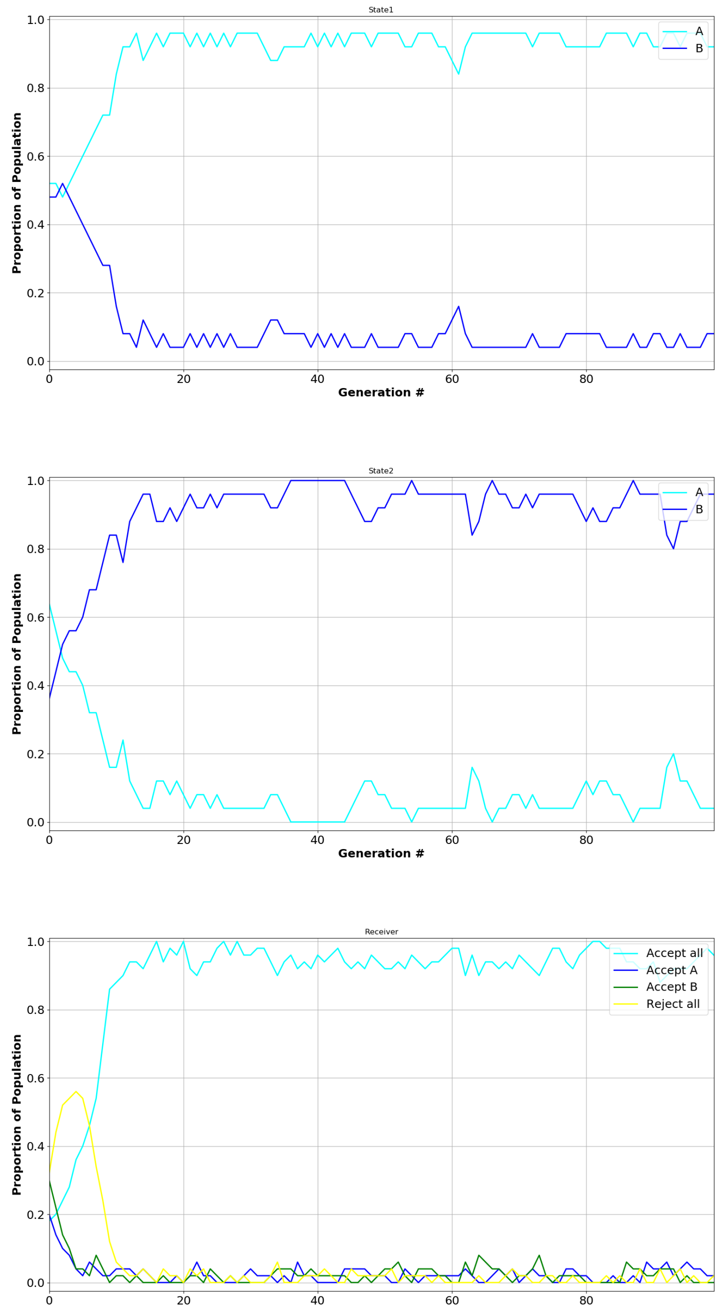

Figure 4: Sample simulation of a 2x2x2 rejection game, $\lambda = 0.5$. The first plot shows the evolution of probabilities for s_1, the second for s_2, the third for the listener. In the end, A signals s_1, and B signals s_2.

4 Rejection Game with Costly Signalling

The plan is now to expand the rejection game to include a system of differential costs, in order to account for the emergence of denial. Note that there is no circularity in assuming rejection to explain denial. Rejection is developmentally and conceptually prior to the expression of negation. While rejection may be linguistically realized ('No!'), it need not be: your dog's refusal to accept an offer of food is as good a rejection as any. Evidence from the cognitive development of infants shows that children acquire the ability to reject (an offer or a prohibition) well within the single-word utterance period, and only later do they acquire negation as a truth-functional device (Tommasello [2003]; Dimroth [2010]).

A key advantage of syntactic complexity is the possibility of overcoming the expressive limitations of a small lexicon (Skyrms [2010]; Kirby *et al.* [2015]). For concreteness, we suppose that the agents' environment is a set $I = \{s_1, s_2, s_3, s_4\}$ of information states, but that the speaker has two signals, m_1 and m_2 (see Figure 5). In this setting, the information states outnumber the available signals, hence the only option for the speaker is to signal with a significant degree of ambiguity.

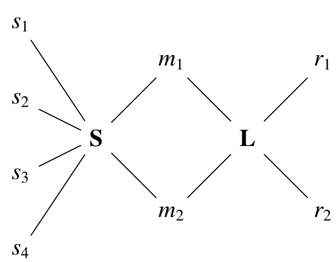

Figure 5: A 4x2x2 rejection game

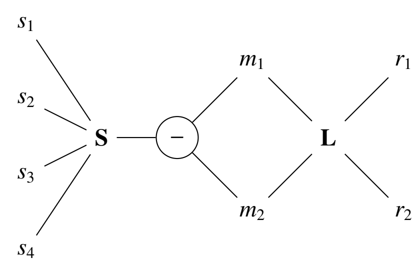

Figure 6: A 4x2x2 rejection game with costly signals. The '−' node indicates an optional production cost.

As we discussed above, we employ both production costs and social costs. As Figure 6 shows, the speaker's signalling possibilities double thanks to production costs, for she can assert not only m_1 and m_2, but also costly counterparts $-m_1$ and $-m_2$. Here '−' represents the additional effort required to produce a phonological and morphological modification of an utterance.

However, resorting to the costly signals $-m_1$ and $-m_2$ is somewhat undesirable for the speaker, who gets higher utilities if cost-free signals are accepted. At an extreme, if the production cost c_p of a signal is too high, the speaker will not even bother to send it, for even the benefits of a successful interaction are not worth shouldering the costs. Simulations show that costly signals are likely to be played, whatever the frequency of the state, roughly if $c_p \leq 0.3 \times a$, where a is the speaker's acceptance payoff (Figure 7). Moreover, when they are used, costly signals tend in the long run to be confined to the less frequent states: this is to be expected, as the speaker uses cheap signals for the states more frequently encountered. For the rest of the paper, we set $c_p = 0.003 \times a$, which is much smaller. This number is arbitrary, but still representative of a generalization.

Before discussing social costs, our next step is to account for compositionality. There are four

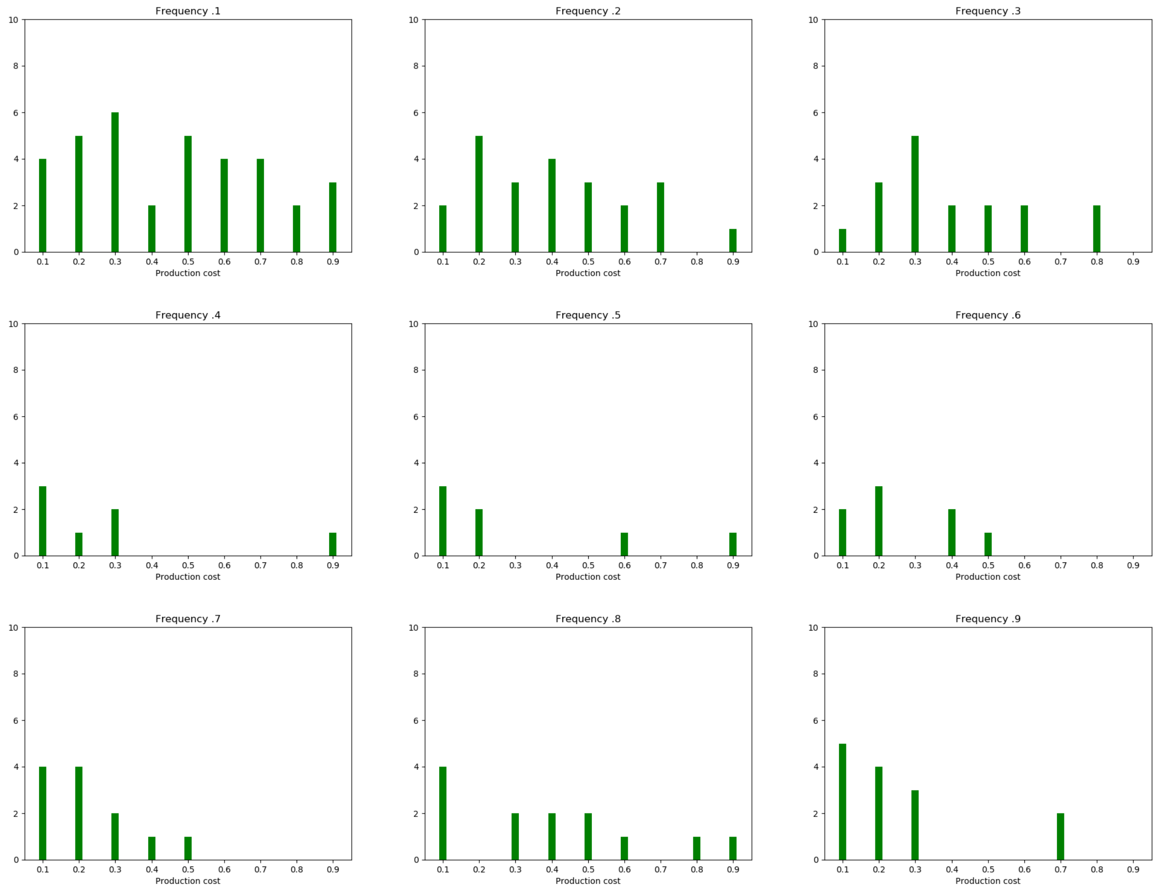

Figure 7: Mean occurrences of marked signals (with $\lambda = 0$) by state frequency from 0.1 to 0.9. Production costs range from $a \times 0.1$ (left) to $a \times 0.9$ (right), where a is the speaker's acceptance payoff. Marked signals are more likely to be produced for less frequent states (top 3 squares) and for lower costs (left columns).

states and, thanks to production costs, four signals, but the signals shouldn't be randomly distributed over the states. We build on the insight that compositionality is 'lifted' from observable relations between information states (Franke [2016]; Steinert-Threlkeld [2016]). A structured information space $J = \langle I, \rho \rangle$ is a set I of information states together with a relation ρ on $I \times I$. For present purposes, we assume states are pairwise related: in other words, $\rho = \{\langle s_1, s_2 \rangle, \langle s_3, s_4 \rangle\}$ (see Figure 8). In this scenario, there are various equally optimal outcomes: for example, the speaker sends m_1 to signal s_2 in one ρ-pair of states, sends m_2 to signal s_3 in another ρ-pair, and sends the marked signals for their correlated states, s_1 and s_4 respectively. In these circumstances ρ splits the difference between signals with a small production cost and signals without.

We model compositionality by modifying the evolutionary dynamics, following Franke ([2016]),

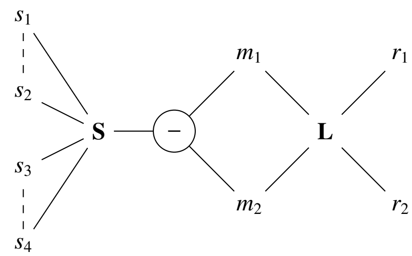

Figure 8: A 4x2x2 rejection game with marked signals in a structured space

by making the development sensitive to the state space metric.[9] By the evolutionary dynamics, acceptance of m in state s, if successful, raises the probability that m will be accepted in s again. We now assume, in addition, that acceptance of m in s, if successful, increases by a small margin the probability that $-m$ will be accepted in the ρ-related state s', and *vice versa*. Rewards for successful actions 'spill over' along the relation ρ on states, for pairs of signals $m, -m$. Thus, ρ codifies the relationship between using m in a state and $-m$ in another.

The upshot of spill-over, concerning the behaviour of the listener, is summarized in Figure 9. Let's suppose for illustration that acceptance of m is successful in s_1. Because of the rule for rejection, m is rejected in every other state at equilibrium (Figure 9, left). Furthermore, let's suppose that s_1 is ρ-related to s_2. Spill-over makes it likely that $-m$ will be accepted in s_2, and therefore (again, in equilibrium) that $-m$ is rejected everywhere but in s_2 by the rule of rejection (Figure 9, right). Spill-over does not consist in assuming that the listener interprets m and $-m$ as mutually exclusive, for ρ need not be the relation of being opposite to. Spill-over assumes that listeners tend to take different but related signals to apply to different but related information states, and allows us to systematically trace the correlation between m and $-m$ to a relation ρ between states.

	s_1	s_2	s_3	s_4
Accept	m			
Reject		m	m	m

	s_1	s_2	s_3	s_4
Accept	m	$-m$		
Reject	$-m$	m	$m, -m$	$m, -m$

Figure 9: Distribution of signals per states given a spill-over dynamics. The table on the left shows the distribution imposed by rejection, the table on the right shows how signals are used in equilibrium if s_1 is ρ-related to s_2.

As far as the agents in the model are concerned, m and $-m$ could be an adjective 'warm' and its superlative 'warmest', a noun 'dog' and its plural 'dogs', a verb 'talk' and its past 'talked': the pairs of signals are related in their form and in their meanings (by some relation ρ), but nothing yet suggests that the morphologically complex element is negative, so to speak. Simulations show,

[9] More precisely, Franke describes a 'spill over' approach to compositionality in the context of a reinforcement learning model, whereas our model of cultural evolution uses a different dynamics for updating probabilities. We maintain the spirit of Franke's approach if not the letter.

moreover, that the space of possible strategies is too unconstrained for speaker and listener to find an optimal solution to the game.

Our next and final step is intended to account for the incompatibility of assertion and denial. We assume that the use of m and $-m$ in the same state receives immediate negative repercussions, in the form of social costs. This way, we trace the logical force of negation to environmental feedback received by the speaker for using a pair m and $-m$ of signals as if they were synonyms. As discussed above, this assumption is called polarity. For any pair $m, -m$ of costless/costly signals:

Polarity Social costs are inflicted if m and $-m$ are sent in the same state.

There is a (marked) difference between m and $-m$, which is perceived overtly, so that any attempt to use m and $-m$ as synonyms triggers a reaction from the social environment: a public fallout, perhaps in the form of a range of disagreeable reactions such as scorn, mistrust, and blame. Incompatibility in the meaning of m and $-m$ is thus traced to the agents' reactive attitudes: we assume that the reactions are there first, and it is from these that the perceived incompatibility in meaning derives.

Polarity allows us to set aside unintended interpretations of m and $-m$ (as adjective and superlative, or noun and plural, and so forth). Moreover, polarity has an effect on the agents' optimization problem. Since a number of strategies (those in which m and $-m$ are sent in the same state) become inefficient for the speaker (so long as, plausibly, social costs outweigh the benefits of acceptance), the search space for equilibrium strategies is much reduced.

We can now calculate the conditions for an optimal weak Nash equilibrium to obtain. A pair of strategies is an equilibium just in case each strategy is a best response for a player, given the other player's strategy. A strategy is a best response for a player if the player has no strategy with higher expected utility given what the other player does. Since costs partly determine a player's choice, the overall value of player i's choice is its benefit eu_i minus its total costs c. Since we have two types of costs—production costs c_p and social costs c_s—we may factor c in two components:

$$w_i(s, m, r) = eu_i(s, m, r) - c_p(s, m) - c_s(s, m)$$

where s is an information state, m a signal, r a reaction, and the expected utility $eu_i(s, m, r)$ is defined as above in terms of strategies. Cost functions for production $c_p(s, m)$, and social fallout $c_s(s, m)$, determine the discounting factor. Costs only influence speaker's behaviour, hence $w_L(s, m, r) = eu_L(s, m, r)$.

We now calculate a simple analytical relation between the value of sending a signal and the two types of costs, following the reasoning in (van Rooij [2003]). For simplicity, consider a rejection game with two states, s_1, s_2, two signals, m_1, m_2, and two reactions, r_1, r_2. An optimal signalling arrangement is one in which S signals unambiguously, and L accepts all signals: for example, S asserts m_1 in s_1 and m_2 in s_2, or else m_2 in s_1 and m_1 in s_2, and L accepts all, for she has no reason to reject (the speaker is always truthful). We only reason through the former case, since the latter is analogous.

Let r_1 be acceptance. In order for S to assert m_1 in s_1 and m_2 in s_2, it must be the case that $w_S(s_1, m_1, r_1) \geq w_S(s_1, m_2, r_1)$ and $w_S(s_2, m_2, r_1) \geq w_S(s_2, m_1, r_1)$. Let $eu_S(s, m, r_1) = a$ for all

s, m, and 0 otherwise. Let's assume that producing m_1 is free, hence $c_p(s_1, m_1) = c_p(s_2, m_1) = 0$, while m_2 has a constant production cost $c_p(s_1, m_2) = c_p(s_2, m_2) = x > 0$. Intuitively, m_1 is the unmarked choice, while m_2 $(= -m_1)$ is marked. If $w_S(s_1, m_1, r_1) \geq w_S(s_1, m_2, r_1)$ and $w_S(s_2, m_2, r_1) \geq w_S(s_2, m_1, r_1)$, then we have

$$a - c_s(s_1, m_1) \geq a - x - c_s(s_1, m_2) \quad \text{and} \quad a - x - c_s(s_2, m_2) \geq a - c_s(s_2, m_1)$$

That is, $c_s(s_1, m_1) < x + c_s(s_1, m_2)$ and $x + c_s(s_2, m_2) < c_s(s_2, m_1)$. Since x is the production cost for a marked signal, we can assume that x is a vanishingly small quantity: the effort of producing an extra morpheme compared to the social fallout of uncovered deception is presumably quite small. So, social costs must be conversely distributed in the different states:

$$c_s(s_1, m_1) < (1 + \epsilon)c_s(s_1, m_2) \quad \text{and} \quad c_s(s_2, m_1) > (1 + \epsilon)c_s(s_2, m_2)$$

So long as these inequalities hold, the speaker prefers to signal unambiguously, hence the listener accepts every signal, and the resulting equilibrium is optimal. It is straightforward to extend the use of production costs to a greater number of states and signals.

It is not feasible to calculate how probabilities evolve for thousands of rounds under the dynamics we have described, without relying on a computer simulation. We have carried out simulation studies of rejection games with costly signalling that show significant results: the agents do find optimal equilibria, with the properties we have described, and the results are robust under reasonable assignments of values to the parameters in the model. An illustration is given in Appendix B.

5 Adequacy of the Account

Our model predicts that a community of signallers who behave as described and conform to the norms indicated is likely to develop a signalling system in which expressions involving a production cost (and hence marked) are accepted just in case their costless (and hence unmarked) counterpart is rejected, over evolutionary time. The account has multiple components:

(i) Stalnakerian speech acts: an update proposal followed by acceptance or rejection, in which the speaker has an interest in avoiding rejection, and the hearer has an interest in accepting unambiguous signals. This is the structure of the overall framework.

(ii) Sympathy: a measure of the degree of cooperativity in the conversation, which contributes to establishing conventional meanings.

(iii) Production costs, attached to some signals, by which we account for the fact that signals to deny are marked.

(iv) Social costs, inflicted upon unveiled deception, by which we account for the semantic difference between signals with a production cost and signals without.

There are three parameters to keep track of, corresponding to (ii), (iii), and (iv). The listener's preferences, codified in her utility function, together with some degree of sympathy (ii), count as an anti-ambiguity tendency against the use of a signal in more than one state. The causal role of production costs is tied to the notion of markedness (iii). Social costs count as an anti-synonymy tendency against the use of a signal m and its counterpart $-m$ in the same state (iv). A consequence of these assumptions is that denial shares certain key properties with negation, discussed in Section 2. We review them in order.

5.1 Markedness

We have assumed that some signals come with a small production cost. This assumption (component (iii) above) leads to a simple game-theoretic account of a first observation we have made concerning denial.

Markedness: Expressions used to deny are marked.

As we saw (Figure 7), even a small production cost induces the speaker to prefer costless signals. Therefore, the use of a costly expression, rather than its cheaper near-synonym (if there is one), tends to be interpreted as signalling a rather infrequent, unexpected, or non-stereotypical event (Horn [1984]). Someone commenting on Fred by using (1b) rather than (1a), may therefore convey that Fred is not fully or stereotypically happy:

(1) a. Fred is happy.
    b. Fred isn't sad.

Interpreting costs as markedness, we have shown that, plausibly, the use of marked signals is confined to less probable states. The correlation between markedness and frequency is well-documented in empirical studies of markedness (Pustet [2009]; Bybee [2011]). We thus capture an aspect of the notion. While markedness in natural language is far more complex, this can be a reasonable first approximation.

5.2 Incompatibility and logical properties

Social costs implement a rigid restriction against the synonymy of m and $-m$. In fact, this restriction captures their incompatibility.

Incompatibility: Denial conveys incompatibility.

Let '$\perp$' symbolize that the speaker's payoffs are multiplied by 0: in a way, this is an extreme proposal for implementing social costs (component (iv) above). Then, with respect to a fixed state, and assuming the listener always accepts, the rule of contradiction holds:

$$\frac{m \quad -m}{\perp}$$

In other words, if m and $-m$ are sent in the same state and the listener always accepts, the speaker will incur a penalty $\perp$. The rule of contradiction obtains as a consequence of polarity: a restriction on the interaction between speaker and listener, which limits the space of possible strategies for the speaker, facilitating the search for equilibria. Absent social costs or if the social cost parameter has too little impact, this search is too complex to result in convergence, other things being equal. In our simulations, we have kept $c_s > a$, that is, social costs outweigh the benefit of acceptance. This condition gives a reasonable lower bound on the value of costs, and it can be justified by considering that, if $a \geq c_s$, ignoring social repercussions could still be all things considered advantageous for the speaker.

Further assumptions reveal related logical properties besides incompatibility. Whatever state the speaker is in, she is in one of two ρ-related states (given the assumptions we made about ρ). Assume, again, that the listener always accepts. By spill-over, acceptance of m in one state leads to acceptance of $-m$ in the other, or *vice versa*. The speaker can then reason as follows: if m is sent, either it is accepted to my benefit, or else its acceptance leads to a penalty, in which case $-m$ is accepted to my benefit (or *vice versa*). Meta-reasoning about the game thus validates the rules of *reductio* R_1 and R_2:

$$\dfrac{\begin{array}{c} m \\ \vdots \\ \perp \end{array}}{-m}\,\mathrm{R}_1 \qquad\qquad \dfrac{\begin{array}{c} -m \\ \vdots \\ \perp \end{array}}{m}\,\mathrm{R}_2$$

This generalization is observed, for example, if there are only two states. In this setting we predict that, from the perspective of the speaker, and assuming the hearer always accepts, '$-$' has the properties of classical negation, '$\neg$'. In particular, '$-$' satisfies the principles laid down by Timothy Smiley ([1996]) for the logic of denial. Accordingly, denial is equivalent to the assertion of a negation:

$$-A \text{ if, and only if, } +\neg A$$

where '$+$' is a force marker for assertion. It is in this sense that our symbol '$-$' captures the basic logical properties of negation.

On the present account, the use of signals is determined by the preferences of the agents and by a system of production and social costs. We have shown that given constraints on conversation that include assertion, rejection, and sanctions, marked signals evolve so that markedness tracks the difference between ρ-related states. This pattern of use does not depend on what ρ actually is. We only have assumed that ρ is a state-to-state binary relation hardwired into the agents' environment, and that it is salient to them. There is, of course, no scarcity of candidate relations between information states in our natural environment, hence it is reasonable to suppose that marked and unmarked signals, whose development is supported by the need to overcome the expressive limitations of a finite lexicon, should be repurposed to track binary relations that we frequently encounter. Examples may be the relation $\{\langle \mathtt{hot}, \mathtt{cold}\rangle, \langle \mathtt{dry}, \mathtt{wet}\rangle\}$ to talk about the weather, or the relation $\{\langle \mathtt{happy}, \mathtt{sad}\rangle, \langle \mathtt{healthy}, \mathtt{sick}\rangle\}$ to talk about a friend. These are instances of contrariety. Gerrymandered relations are not plausible candidates, simply because human beings are hardly ever in an information space in which gerrymandered relations tend to be salient. We have to imagine that games with the structure we described are played over and over in the course of human history. Even if some relation other than contrariety were observed in the environment, it would have to be observed frequently enough to polarize the use of a pair of marked/unmarked signals over evolutionary time.

We make the hypothesis that marked and unmarked signals, whose development we accounted for in a social setting, may find a use in tracking contrariety. This relation is, of course, very salient to us, and as noticeable as the difference between hot and cold weather. It's plausible therefore that the use of polar signals may be appropriated to track contrariety.

5.3 Compositionality

Finally, we assumed a semantic characterization of $-m$ as a signal whose meaning is a function of the meaning of m.

Compositionality: Expressions used to deny are semantically compositional.

In equilibrium conditions the meaning of $-m$ is whatever state is ρ-related to the state that counts as the meaning of m. This is a functional relation in the semantics, which is enforced in the long run by the spill-over distribution of reinforced probabilities.

There are other approaches to compositionality in signalling games. In a Skyrms-Barrett game there are two senders, or equivalently a single sender who plays twice (Barrett ([2009]); Skyrms ([2010]); LaCroix ([forthcoming])): multiple signals combined together may express complex meanings. However, some misgivings about Skyrms-Barrett games have been raised by Franke ([2016]), who complains that the component signals are not independently meaningful to the agents.

A related proposal is due to Steinert-Threlkeld ([2016]). His signalling game for negation can be understood as a Skyrms-Barrett game, in which an optional first signal '$\neg$' modifies the interpretation of the second signal, m_1 or m_2. States s_1, s_2, t_1, t_2, are endowed with a binary relation so that for example, $s_1 \sim s_2$ and $t_1 \sim t_2$. The intended interpretation is that s_2 and t_2 are the opposites of s_1 and t_1, respectively. Updates are such that, if m_1 correlates with s_1, then $\neg m_1$ correlates with s_2, likewise for t_1 and t_2. While we share some aspects of this approach, we find that too much is left to interpretation: nothing in the game indicates that m and $\neg m$ are incompatible. For example, let the four states be two pairs of twins, and the relation $\sim$ hold for siblings. The game would then show that two names m_1 and m_2 suffice for four people, pairwise related, thanks to a sign '$\neg$' standing for 'twin of'.

Our account is inspired by Franke ([2016]), who introduces a similarity relation between states and signals, and spill-over: an update rule for probabilities that is sensitive to such relations (O'Connor [2014]). In Franke's framework, as well as in ours, signals are syntactically atomic, and compositionality is accounted for as a relation in the semantics. We acknowledge that the lack of syntactic sophistication is a limit of our account, and indicate this limitation as a topic for further research.

6 Conclusion

In this paper, we have considered the question why all human languages have a device to express denial. On our proposal, a form of 'proto-negation' emerges in gradual steps, driven by selectional forces operating on conversations. We have isolated some plausible and relatively undemanding constraints on behaviour, and shown that a signalling device that resembles negation (with respect to markedness, incompatibility, and compositionality) evolves under a suitable dynamics.

By necessity, our model is an idealisation which represents a simplification of complex historical and linguistic phenomena. However, the abstract nature of the evolutionary model employed is also a strength, in that it can be seen as a description of the general conditions under which we might expect negation to develop, and apply to many different particular historical dynamics.

Our hypothesis is that denial is an entrenched behavioural regularity. We have assumed that costly signals may spontaneously appear within a population. Their use is then supported by the expressive needs of an informationally rich environment confronted with a small-size lexicon, and constrained by social costs. Production costs account for markedness, and social costs account for incompatibility. Finally, we hypothesised how such device can be repurposed to describe relations of contrariety that are easily observable in our environment. The proposed account is based on a fundamentally Darwinian understanding of evolutionary development (Croft [2000]), formalized by the Wright-Fisher process (Ewens [1979]). Accordingly, individual-level selection between agents playing a rejection game is what best explains the universality of negation in human languages.

Appendix A: The Wright-Fisher Model with Selection

We use populations to model the behaviour of individuals. A population is a set of types, each of which instantiates a pair of strategies, one for speaker and one for listener. Let there be two types, τ_i and τ_j, whose strategies are $\langle \sigma_i, \sigma'_i \rangle$ and $\langle \sigma_j, \sigma'_j \rangle$ respectively. If all, or nearly all, members of a population P are type τ_i, the individual whose behaviour is modelled by P behaves always, or nearly always, by σ_i as speaker and by σ'_i as listener. If there are about as many τ_i-members of P as there are τ_j-members, the individual plays σ_i or σ_j as speaker with about a 50% chance, likewise as listener. This setup allows us to import methods of population genetics into the study of linguistic behaviour.

We assume that the distribution of types is initialized at random, and by the Wright-Fisher equation we calculate the probability that a strategy played with probability p at time $t-1$ will be played with probability p' at time t. The Wright-Fisher equation describes a discrete time evolutionary process.

We take a population of types of size N for each state. Suppose that, as speaker, an agent could be in one of two states s_1 or s_2. Suppose we are in state s_1. Consider a type τ_i, defined by the strategies available in s_1. Let x_t be the number of elements of s_1's population that are τ_i at time t. The probability that the speaker in state s_1 plays by τ_i at time t is simply the proportion of

τ_i-elements in that state:

$$p = \frac{x_t}{N}$$

Let $x_t = z$, and let the range of $\bar{z}$ be $\{0, 1, ..., N\}$. The Wright-Fisher equation (first, without selection) gives the transition probabilities from a time to the next: in other words, we calculate the likelihood that the number of τ_i-elements at $t + 1$ will be $\bar{z}$, given that that number is z at t:

$$Pr(x_{t+1} = \bar{z}|x_t = z) = \frac{N!}{\bar{z}!(N-\bar{z})!} p^{\bar{z}}(1-p)^{N-\bar{z}} \tag{WF}$$

where $N!/\bar{z}!(N - \bar{z})!$ is the binomial coefficient of the probability mass function. This is a basic model of drift, in which we assume that at each time, probabilities are determined by sampling with replacement from the population at the previous time. (Intuitively, at any time, each element of the population looks back at the population at previous time and picks a type at random to imitate.) The next step is to add selection.

The basic concept is familiar from biology: types are selected that have higher fitness. In evolutionary theory, the fitness of a strategy is defined as the number of offspring of a population element that plays that strategy. This definition is simple and domain-general, and can be applied to linguistics: linguistic behaviour is selected that tends to outperform alternatives in communication, maximizing informativity and minimizing costs (if there are any). Let $u(i, j)$ be the average expected utility of τ_i playing against τ_j. Assuming that each agents is speaker or hearer about half the time, this is simply calculated as follows (van Rooij [2008]):

$$u(i, j) = \frac{1}{2}u_S(\sigma_i, \sigma'_j) + \frac{1}{2}u_L(\sigma_j, \sigma'_i)$$

The relative fitness of types τ_i and τ_j, assuming for simplicity that these are the only two types in the population, are defined as follows:

$$f_i = p \times u(i, i) + (1 - p) \times u(i, j) \qquad f_j = (1 - p) \times u(j, i) + p \times u(j, j)$$

Typically, different combinations of strategies will have different fitness, which can be the basis for differential selection. We define the gap between the strategies' fitness as a selection coefficient s:

$$f_i : f_j = 1 : 1 - s$$

If $s > 0$, τ_i has an advantage over τ_j, and if $s < 0$, a disadvantage. If $s = 0$, there is no selection and all evolutionary causal forces remaining are oscillations due to drift. Probabilities of types with selection are given by the formula:

$$p^* = \frac{x_t(1 - s)}{x_t(1 - s) + N - x_t}$$

and finally p^* is written into the transition probabilities equation WF instead of p. The simulation studies of the rejection game discussed below were conducted under the Wright-Fisher model with selection.

Appendix B: Sample Simulation

Figures 10 and 11 show the results of two sample simulations.[10] In both simulations, we stipulated that s_1 and s_3 are ρ-related, and so are s_2 and s_4. We ran the first simulation for payoff $a = 3$, sympathy $\lambda = 0.5$, production costs $c_p = 0.1$, social costs $c_s = 5$. This parameter setting is consistent with the assignment of values we argued for in the paper: $c_s > a$ and $c_p = a/30$. The behaviour of the speaker in this setting is represented by Figure 10 in states s_1, s_2, and s_3. Observe that the speaker behaves more or less randomly at the beginning, and eventually (well within the space of 500 generations, or rounds of interaction) she stabilizes around sending $-A$ in s_1, $+B$ in s_2, and $+A$ in s_3, each with about .8 chance. Signalling probabilities for $-A$ in s_1 are slightly lower than those for positive signals $+A$ and $+B$, and this is essentially due to the speaker's reluctance to using costly signals. (There is a fourth state s_4, out of picture, in which the speaker sends $-B$ with roughly the same probability.) Due to the inherent randomness of the Wright-Fisher dynamics, probabilities oscillate without stabilizing. For values of λ approximating 1, keeping the remaining parameters the same, the probability that a speaker uses perfectly unambiguous signalling strategies gets closer to 1.

The behaviour of the listener for the same simulation is plotted in the top box of Figure 11. Note that the listener is initially inclined to reject everything, as the speaker behaves randomly, and eventually accepts nearly everything (with about 0.9 chance) as the speaker's behaviour becomes more orderly. In the final plot (bottom of Figure 11), we illustrate the speaker's behaviour with sympathy $\lambda = 0.99$ but social costs $c_s = 0$: all cooperativity but no social norms. The plot illustrates the importance of social costs for the development of negation: even stipulating that cooperativity is fully rewarding, absence of social costs leaves too much noise for the agents to optimize. In this simulation, the listener ends up rejecting nearly everything (out of picture).

[10] Code for this work can be found at `https://github.com/gsbardolini/evolutionofdenial`.

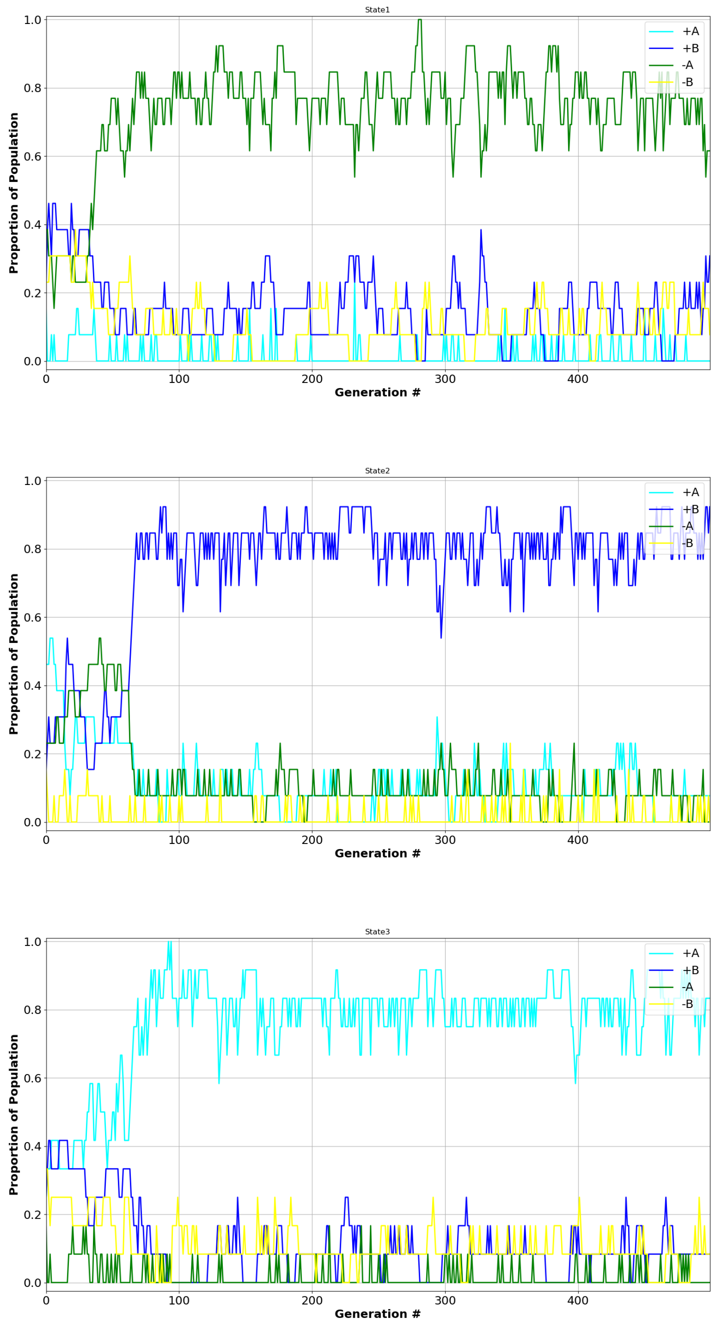

Figure 10: Sample simulation of a 4x2x2 rejection game with marked signals in a structured space with $\lambda = 0.5$, same simulation as Figure 11. The plots represent, from top, the behaviour of the speaker in states 1, 2, and 3. (There is a fourth state, out of picture.) Note that the speaker behaves more or less randomly for about the first 100 rounds of the game, and eventually signals with probabilities around 0.8.

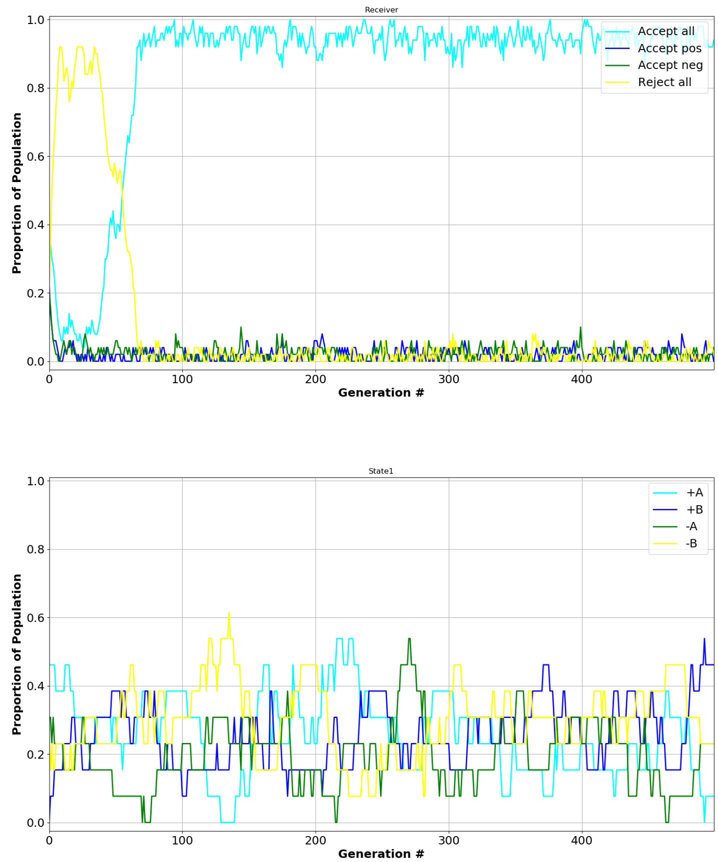

Figure 11: The top figure represent the behaviour of the listener, in the same simulation as Figure 10. Note that the listener initially rejects all, but as soon as the speaker begins to sort signals for states, she ends up accepting. For comparison, the bottom figure represents the speaker's behaviour in a state (in this case, `state1`) if we set very high sympathy $\lambda = 0.99$ but no social costs $c_s = 0$. The result is that the listener fails to enforce honest signalling and indeed winds up rejecting everything (out of picture).

Acknowledgments

We would like to thank the Editors and two referees for this journal for their helpful and insightful comments. This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 758540) within the project *From the Expression of Disagreement to New Foundations for Expressivist Semantics*.

Luca Incurvati

Institute for Logic, Language and Computation
University of Amsterdam
Amsterdam, The Netherlands
l.incurvati@uva.nl

Giorgio Sbardolini
Institute for Logic, Language and Computation
University of Amsterdam
Amsterdam, The Netherlands
g.sbardolini@uva.nl

References

Barrett, J. [2009]: ‘The Evolution of Coding in Signaling Games’, *Theory and Decision*, **67**, pp. 223–237.

Barrett, J. and Zollman, K. [2009]: ‘The Role of Forgetting in the Evolution and Learning of Language’, *Journal of Experimental and Theoretical Artificial Intelligence*, **21**, pp. 293–309.

Batali, J. [1998]: ‘Computational Simulations of the Emergence of Grammar’, in J. R. Hurford, M. Studdert-Kennedy and C. Knight (*eds*), *Approaches to the Evolution of Language: Social and Cognitive Bases*, Cambridge: Cambridge University Press, pp. 405–426.

Berto, F. and Restall, G. [2019]: ‘Negation on the Australian Plan’, *Journal of Philosophical Logic*, **1**, pp. 1–26.

Bloomfield, L. and Hockett, C. F. [1984]: *Language*, Chicago: University of Chicago Press.

Blythe, R. and Croft, W. [2012]: ‘S-curves and the Mechanisms of Propagation in Language Change’, *Language*, **88**, pp. 269–304.

Bond, O. [2012]: ‘A Base for Canonical Negation’, in D. Brown, M. Chumakina and G. Corbet (*eds*), *Canonical Morphology and Syntax*, Oxford: Oxford University Press, pp. 20–47.

Boyd, R. and Richerson, P. [1985]: *Culture and the Evolutionary Process*, Chicago: University of Chicago Press.

Bybee, J. [2011]: ‘Markedness: Iconicity, Economy, and Frequency’, in J.-J. Song (*ed.*), *The Oxford Handbook of Linguistic Typology*, Oxford: Oxford University Press, pp. 131–147.

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

Crawford, V. P. and Sobel, J. [1982]: ‘Strategic Information Transmission’, *Econometrica*, **50**, pp. 1431–1451.

Croft, W. [2000]: *Explaining Language Change: An Evolutionary Approach*, London: Longman.

Dahl, Ö. [1979]: 'Typology of Sentence Negation', *Linguistics*, **17**, pp. 79–106.

Dahl, Ö. [2010]: 'Typology of Negation', in L. Horn (*ed.*), *The Expression of Negation*, Berlin/New York, NY: Mouton De Gruyter, pp. 9–38.

De Swart, H. [2010]: *Expression and Interpretation of Negation: An OT Typology*, Dordrecht: Springer.

Dimroth, C. [2010]: 'The acquisition of negation', in L. Horn (*ed.*), *The expression of negation*, Berlin/New York, NY: Mouton De Gruyter, pp. 39–72.

Ewens, W. [1979]: *Mathematical Population Genetics*, Berlin: Springer.

Farrell, J. and Rabin, M. [1996]: 'Cheap Talk', *Journal of Economic Perspectives*, **10**, pp. 103–118.

Frana, I. and Rawlins, K. [2019]: 'Attitudes in discourse: Italian polar questions and the particle *mica*', *Semantics and Pragmatics*, **12**, pp. 1–48.

Franke, M. [2016]: 'The evolution of compositionality in signaling games', *Journal of Logic, Language, and Information*, **25**, pp. 355–377.

Grice, H. P. [1975]: 'Logic and Conversation', in M. Ezcurdia and R. J. Stainton (*eds*), *The Semantics-Pragmatics Boundary in Philosophy*, Peterborough, Ontario: Broadview Press, pp. 47–59.

Grice, H. P. [1989]: *Studies in the Way of Words*, Cambridge, MA: Harvard University Press.

Haldane, J. B. S. [1992]: 'Animal communication and the origin of human language', *Current Science*, **63**, pp. 604–611.

Hauser, M. D., Yang, C., Berwick, R. C., Tattersall, I., Ryan, M. J., Watumull, J., Chomsky, N. and Lewontin, R. C. [2014]: 'The mystery of language evolution', *Frontiers in Psychology*, **5**, pp. 401.

Herrnstein, R. J. [1970]: 'On the law of effect', *Journal of the Experimental Analysis of Behavior*, **13**, pp. 243–266.

Hofbauer, J. and Sigmund, K. [1998]: *Evolutionary Games and Population Dynamics*, Cambridge: Cambridge University Press.

Hollén, L. and Radford, A. [2009]: 'The development of alarm call behaviour in mammals and birds', *Animal Behaviour*, **78**, pp. 791–800.

Horn, L. [1984]: ‘Toward a New Taxonomy for Pragmatic Inference: Q-Based and R-Based Implicature’, in D. Schriffrin (*ed.*), *Meaning, Form, and Use in Context: Linguistic Applications*, Washington, DC: Georgetown University Press, pp. 11–42.

Horn, L. [1989]: *A Natural History of Negation*, Chicago: University of Chicago Press.

Huttegger, S. [2007]: ‘Evolution and the Explanation of Meaning’, *Philosophy of Science*, **74**, pp. 1–27.

Huttegger, S. and Zollman, K. [2011]: ‘Signaling Games’, in A. Benz, C. Ebert, G. Jäger and R. van Rooij (*eds*), *Language, Games, and Evolution: Trends in Current Research on Language and Game Theory*, Berlin: Springer, pp. 160–176.

Huttegger, S. M., Skyrms, B., Smead, R. and Zollman, K. J. S. [2010]: ‘Evolutionary Dynamics of Lewis Signaling Games: Signaling Systems vs. Partial Pooling’, *Synthese*, **172**, pp. 177–191.

Incurvati, L. and Sbardolini, G. [unpublished]: ‘The Rejection Game’, unpublished.

Incurvati, L. and Schlöder, J. J. [2017]: ‘Weak Rejection’, *Australasian Journal of Philosophy*, **95**, pp. 741–760.

Jackendoff, R. [1999]: ‘Possible Stages in the Evolution of the Language Capacity’, *Trends in Cognitive Sciences*, **3**, pp. 272–279.

Jespersen, O. [1917]: *Negation in English and Other Languages*, Copenhagen: A. F. Høst.

Kamenica, E. and Gentzkow, M. [2011]: ‘Bayesian Persuasion’, *American Economic Review*, **101**, pp. 2590–2615.

Kean, M.-L. [1992]: ‘Markedness’, in W. Bright (*ed.*), *International Encyclopedia of Linguistics*, Oxford: Oxford University Press, pp. 390–391.

Kirby, S., Tamariz, M., Cornish, H. and Smith, K. [2015]: ‘Compression and Communication in the Cultural Evolution of Linguistic Structure’, *Cognition*, **141**, pp. 87–102.

Lachmann, M., Számadó, S. and Bergstrom, C. T. [2001]: ‘Cost and conflict in animal signals and human language’, *Proceedings of the National Academy of Sciences*, **98**, pp. 13189–13194.

LaCroix, T. [forthcoming]: ‘Using Logic to Evolve More Logic: Composing Logical Operators via Self-Assembly’, *The British Journal for the Philosophy of Science*.

Lassiter, D. [2008]: ‘Semantic Externalism, Language Variation, and Sociolinguistic Accommodation’, *Mind and Language*, **23**, pp. 607–633.

Levinson, S. C. [2000]: *Presumptive Meanings: The Theory of Generalized Conversational Implicature*, MIT Press.

Lewis, D. [1969]: *Convention*, Cambridge MA: Harvard University Press.

Manin, D. [2008]: 'Zipf's Law and Avoidance of Excessive Synonymy', *Cognitive Science*, **32**, pp. 1075–1098.

Martinez, M. and Godfrey-Smith, P. [2016]: 'Common Interests and Signaling Games: A Dynamic Analysis', *Philosophy of Science*, **83**, pp. 371–392.

Maynard Smith, J. [1991]: 'Honest Signalling: The Philip Sidney Game', *Animal Behaviour*, **42**, pp. 1034–1035.

Maynard Smith, J. and Harper, D. G. C. [1988]: 'The Evolution of Aggression: Can Selection Generate Variability?', *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, **319**, pp. 557–570.

Negroni, M. M. G. [2017]: 'Metalinguistic Negation and Rejection Discourse Markers in Spanish', in M. Roitman (*ed.*), *The Pragmatics of Negation*, Amsterdam: John Benjamins, pp. 17–44.

Nowak, M. A. and Krakauer, D. C. [1999]: 'The Evolution of Language', *Proceedings of the National Academy of Sciences*, **96**, pp. 8028–8033.

O'Connor, C. [2014]: 'The Evolution of Vagueness', *Erkenntnis*, **79**, pp. 707–727.

Osborne, M. J. and Rubinstein, A. [1994]: *A Course in Game Theory*, Cambridge, MA: MIT Press.

Pagin, P. and Westerståhl, D. [2010]: 'Compositionality I: Definitions and Variants', *Philosophy Compass*, **5**, pp. 250–264.

Parry, M. [2013]: 'Negation in the History of Italo-Romance', in D. Willis, C. Lucas and A. Breitbarth (*eds*), *The History of Negation in the languages of Europe and the Mediterranean*, Oxford: Oxford University Press, pp. 77–118.

Pawlowitsch, C. [2008]: 'Why Evolution Does Not Always Lead to an Optimal Signaling System', *Games and Economic Behavior*, **63**, pp. 203–226.

Payne, J. [1985]: 'Negation', *Language Typology and Syntactic Description*, **1**, pp. 197–242.

Penn, D. J. and Számadó, S. [2019]: 'The Handicap Principle: How an Erroneous Hypothesis Became a Scientific Principle', *Biological Reviews*, **95**, pp. 267–290.

Piantadosi, S., Tily, H. and Gibson, E. [2012]: 'The Communicative Function of Ambiguity in Language', *Cognition*, **122**, pp. 280–291.

Pinker, S. and Bloom, P. [1990]: 'Natural Selection and Natural Language', *Behavioral and Brain Sciences*, **13**, pp. 707–784.

Pinker, S. and Jackendoff, R. [2005]: 'The Faculty of Language: What's Special About It?', *Cognition*, **95**, pp. 201–236.

Price, H. [1990]: 'Why 'Not'?', *Mind*, **99**, pp. 221–238.

Pustet, R. [2009]: 'On the Correlation between Discourse Frequency and Structural Complexity in Markedness Theory', *Discours*, **5**, pp. 1–22.

Rumfitt, I. [2000]: '"Yes and" "No"', *Mind*, **109**, pp. 781–823.

Sally, D. [2000]: 'A General Theory of Sympathy, Mind-Reading, and Social Interaction, with an Application to the Prisoners' Dilemma', *Social Science Information*, **39**, pp. 567–634.

Sally, D. [2003]: 'Risky Speech: Behavioral Game Theory and Pragmatics', *Journal of Pragmatics*, **35**, pp. 1223–1245.

Scott-Phillips, T. [2015]: *Speaking our Minds: Why Human Communication is Different, and How Language Evolved to Make it Special*, London: Red Globe Press.

Seyfarth, R., Cheney, D. and Marler, P. [1980]: 'Monkey Responses to Three Different Alarm Calls: Evidence of Predator Classification and Semantic Communication', *Science*, **210**, pp. 801–803.

Skyrms, B. [2010]: *Signals: Evolution, Learning, and Information*, New York: Oxford University Press.

Smiley, T. [1996]: 'Rejection', *Analysis*, **56**, pp. 1–9.

Spence, M. [1973]: 'Job Market Signaling', *The Quarterly Journal of Economics*, **87**, pp. 355–374.

Sperber, D. and Wilson, D. [1995]: *Relevance: Communication and Cognition*, Oxford: Blackwell.

Spike, M., Stadler, K., Kirby, S. and Smith, K. [2017]: 'Minimal Requirements for the Emergence of Learned Signaling', *Cognitive Science*, **41**, pp. 623–658.

Stalnaker, R. [1978]: 'Assertion', *Syntax and Semantics*, **9**, pp. 315–332.

Steels, L. [2011]: 'Modeling the Cultural Evolution of Language', *Physics of Life Review*, **8**, pp. 339–356.

Steinert-Threlkeld, S. [2016]: 'Compositional Signaling in a Complex World', *Journal of Logic, Language and Information*, **25**, pp. 379–397.

Szabó, Z. G. [2000]: 'Compositionality as Supervenience', *Linguistics and Philosophy*, **23**, pp. 475–505.

Taylor, P. D. and Jonker, L. B. [1978]: 'Evolutionary Stable Strategies and Game Dynamics', *Mathematical Biosciences*, **40**, pp. 145–156.

Tommasello, M. [2003]: *Constructing a Language: A Usage-Based Theory of Language Acquisition*, Cambridge, MA: Harvard University Press.

van der Auwera, J. [2009]: 'The Jespersen Cycles', in E. van Gelderen (*ed.*), *Cyclical Change*, Amsterdam: John Benjamins, pp. 35–71.

van Rooij, R. [2003]: 'Being Polite is a Handicap: Towards a Game Theoretical Analysis of Polite Linguistic Behavior', in *Proceedings of the 9th Conference on Theoretical Aspects of Rationality and Knowledge*, TARK '03, pp. 45–58.

van Rooij, R. [2008]: 'Evolutionary Motivations for Semantic Universals', in R. Eckardt, G. Jäger and T. Veenstra (*eds*), *Variation, Selection, Development: Probing the Evolutionary Model of Language Change*, vol. 197 of *Trends in Linguistics*, Berlin: Mouton de Gruyter, pp. 103–142.

Veblen, T. [1899]: *The Theory of the Leisure Class*, New York, NY: Dover Publications.

Weibull, J. [1995]: *Evolutionary Game Theory*, Cambridge, MA: MIT University Press.

Wheeler, B. C. [2009]: 'Monkeys Crying Wolf? Tufted Capuchin Monkeys Use Anti-Predator Calls to Usurp Resources from Conspecifics', *Proceedings of the Royal Society B: Biological Sciences*, **276**, pp. 3013–3018.

Zahavi, A. [1975]: 'Mate Selection—A Selection for a Handicap', *Journal of Theoretical Biology*, **53**, pp. 205–214.

Zanuttini, R. [1997]: *Negation and Clausal Structure: A Comparative Study of Romance Languages*, Oxford: Oxford University Press.

Zipf, G. K. [1949]: *Human Behavior and the Principle of Least Effort*, Oxford: Addison-Wesley Press.

Zollman, K. J. S., Bergstrom, C. T. and Huttegger, S. M. [2013]: 'Between Cheap and Costly Signals: The Evolution of Partially Honest Communication', *Proceedings of the Royal Society B: Biological Sciences*, **280**, pp. 1–8.

Zuberbühler, K. [2002]: 'A Syntactic Rule in Forest Monkey Communication', *Animal Behaviour*, **63**, pp. 293–299.