Games, Norms, and Utterances

MIHAELA POPA-WYATT AND JEREMY L. WYATT

Abstract

A body of work proposes that social-norm change can be explained in terms of game theory. These game theoretic models, however, don't fully account for how and why utterances are used to change social norms. This paper describes the problem and some of the solution elements. There are three existing, relevant, game-based models. The first is a game theoretic model of social norm change (Bicchieri, 2005, 2016). This accounts for how individuals make decisions to adhere to or violate norms, based on empirical expectations of how others will behave. The second is the idea of a conversational game (Lewis, 1979) and its extensions. This posits that speech acts are accommodated in a conversation to make what is said correct play. This feature can explain how some speech acts, such as slurring utterances, change the dynamics of a conversation. The third is a theory of pragmatic inference, known as Rational Speech Act theory (Goodman and Frank, 2016). This is a computational theory of pragmatics, of how listeners interpret utterances and how speakers construct utterances that can be understood. This paper proposes, without setting out the full formal model, that elements of these three theories need to be incorporated together into a game theoretic model of how utterances change long-term social norms.

1. Introduction

All social activities are governed by social norms – informal rules which guide our behaviour. Social norms are not static but change over time. The mechanisms of social norm change have been extensively studied. As part of this, a substantial body of work models social norms using game theory. There are, for example, game theoretic models which account for social norm adherence, violation, and change. These account for a variety of evidence concerning human behaviour. One type of social activity is dialogue. We engage in dialogue to entertain, to inform, and to achieve individual or social goals. Because it is a social activity, dialogue is therefore also a norm-governed activity. There are numerous theories of dialogue, including those that are philosophical, computational, and linguistic in their roots. A number of these theories are also game theoretic. In such models, the atomic move in the dialogue game is an utterance.

doi:10.1017/S1358246124000055

Separately, we know that utterances, whether taken on their own or as part of a larger structure, are used to alter social norms. A good example of social norms evolving through public dialogue is the manner in which social and political speech evolves over time to become, for example, more or less inclusive, racist, or sexist. But dialogue does not just alter norms of speech, it has an effect on the social norms governing other kinds of behaviour. For example, political speech that dehumanizes a group often precedes a sustained campaign of physical violence against that group (Tirrell, 2012).

The problem addressed here is how utterances bring about such changes in social norms. There are two questions. First, why does a speaker make an utterance that (seeks to) alter a social norm? Second, how does that utterance contribute to altering the social norms that apply in a situation? The paper can't provide a complete answer. The paper will, however, describe which elements from three different theories should be combined to provide an answer. It will also set out some features that the combined theory must have.

As mentioned above, the main theoretical tool we will employ is game theory. There are two distinct areas of application of game theory that are relevant. One is social, another is linguistic. In social modelling, game theory has been used to explain a variety of social phenomena: social norm change (Bicchieri, 2005; Bicchieri and Mercier, 2014), the fair distribution of resources (Binmore, 1994b, 1994a; Sterelny, 2021), the emergence of oppression (O'Connor, 2019), and the emergence of social contracts through signalling (Skyrms, 2002). On the side of language, the idea of conversation as a game has a long standing in the philosophy of language, having been first introduced by David Lewis (1979). Moreover, game theoretic models have been used to model pragmatic inference in the form of the theory of Rational Speech Acts (Goodman and Frank, 2016). We can provide initial answers to our two questions by drawing on three of the above: game theoretic models of social norm change; the theory of conversational games; and the theory of Rational Speech Acts. In particular, this paper identifies requirements for a combined theory that can answer the questions.

The remainder of the paper is structured as follows. First, we give some example utterances that illustrate how speech can alter social norms. Second, we describe the idea of conversational games and recent steps to extend this theory to model social norm change. Third, we describe the essence of Bicchieri's game theoretic model of social norm change. In each section, we will identify gaps in the theories. Fourth, we describe the relevant ideas from Rational Speech Act theory. Finally, we sketch how we might fill in the gaps left over.

2. Data

As mentioned above, utterances have an interesting property, which is that they can be used to do things as well as convey information. Take the following utterance:

(1) '*The fake news media isn't my enemy, it's the enemy of the American people*.'

This was a Tweet by the then US President, Donald Trump. The intent of the utterance is clear. The aim was to sow distrust in, and animosity toward, sections of the media in the minds of both undecided and already supportive audience members. In this way, the utterance has the goal to disable that section of the media as an effective public voice and scrutineer of his presidency.

A second example is a dialogue excerpt from the film 'In the Heat of the Night'.

(2) **Gillespie:** '*And just what do you do up there in little old Pennsylvania to earn that kind of money?*'
Tibbs: '*I'm a police officer*.'

Prior to this utterance, Tibbs, who has been arrested on suspicion of murder, is being interrogated by the local police chief, Gillespie. Gillespie does not know that Tibbs is, in fact, a homicide detective. In this excerpt, Gillespie enquires as to how Tibbs came to have more than two hundred dollars in his wallet. The key utterance is the line from Tibbs stating that he is a police officer. This utterance equalizes the power status of the two men, which began as unequal, Gillespie having used the racist derogative 'boy'. It ends later with Gillespie addressing Tibbs with the honorific 'officer'.

On the face of it, these utterances appear to be very different. What is being done, however, at some level of abstraction, is related. Specifically, each utterance seeks to change the rules of the conversation, so as to disempower or re-empower a participant. Both utterances, in context, contributed to changing social norms. Trump's repeated attacks on the press have reduced trust in the media among Republican voters.[1]

[1] The Press Gazette reported a survey by Gallup showing that the percentage of Republicans who trusted the media fell from 32% in 2015 to 10% in 2020. See Majid (2022) https://pressgazette.co.uk/media-audience-and-business-data/trump-vs-media-freedom-of-press-distrust/, and assaults on journalists reached a high level around the 2020 US election, see

In contrast, the utterance in the film 'In the Heat of the Night' was one of several scenes where a black character was established as having equal power to the white characters. The Mr Tibbs character, with this line, makes himself the equal of the Chief Gillespie character. This culminates in a scene where the character of Mr Tibbs returns a face slap given by a powerful and racist white man. This has been called 'the slap heard around the world'. It was reportedly met with cheers (from many black audience members) and shocked cries of 'Oh!'(from many white audience members) in film theatres in the USA on the film's release and is regarded as a landmark social moment. So, the utterance not only changed the conversational dynamics in the scene but was part of a portrayal that changed social norms in America.

In the next section, we will introduce work on conversational games that goes some way to explaining what is happening within a conversation where such a dialogue move is being made. However, we will also argue that this framework cannot alone explain how the utterances cause long-term social change.

3. Conversational Games

Utterances are not stand-alone entities. They are sequenced to form monologues or conversations. From one perspective, an entire social life is simply a sequence of conversations. For our purpose, it is important to emphasize that each utterance influences subsequent utterances in the conversation and each conversation has the capacity to influence future conversations, and future social interactions. We also noted that conversations are norm-governed activities, just like other social activities. We now consider some of the effects that utterances have. This first entails summarising two independent ideas: speech acts and conversational games. We will then combine them to explain the effects of interest.

It is a well-known property of utterances that they don't merely convey information. They can also be used to perform actions or speech acts (Austin, 1975). An example of a speech act is a *performative* such as the utterance 'I now pronounce you man and wife'. This alters the world by creating a binding contract of marriage between two people. Other examples include 'We find the defendant guilty',

https://pressfreedomtracker.us/, where tracking started in 2017 (last accessed on 28 December 2023).

'I name this ship the Queen Elizabeth', and 'I bet you five dollars that it will rain tomorrow'. In each case, the world has changed and there are different norms that apply after each utterance has been made. So, certain utterances can directly alter the social norms which apply in a given context.

Turning from stand-alone utterances to full conversations, an entirely independent observation is that conversations have similarities to games. The term *conversational game* is due to David Lewis (1979), who pointed out that, similarly to a baseball game, a conversation has a score and a scoreboard. This is a way to keep track of the moves in the game and their consequences for the state of play. Each time an utterance is made, the conversational score is updated. Lewis identified a peculiar feature of conversational games, which is that the score updates to make what is said correct play. For example, if I say 'I took my dog for a walk this morning', it puts onto the conversational score, via presupposition, the new information that I have a dog. This is accommodated as correct play and it would be thus inappropriate to ask me later if I have a dog. Another feature of the conversational game is that it evolves according to the rules or norms of conversation. Which conversational norms apply in a particular conversation depends on the participants and the social context in which the conversation takes place. For example, the norms of polite conversation will be different if meeting a VIP than meeting a friend in the pub.

We can usefully combine these two ideas: speech acts and conversational games, to account for the fact that speakers can change the social norms that are salient to a conversation. Mary Kate McGowan (2004) proposed the idea of a particular type of speech act, called a '*conversational exercitive*'. A conversational exercitive is a particular utterance which updates the conversational score so that new norms apply. It changes the permissibility facts in ways that may go unacknowledged by the participants in a conversation. In the example from 'In the Heat of the Night', when Mr Tibbs says 'I'm a police officer', he makes salient the social norms according to which persons of particular professional standing address one another. These social norms now guide the conversation, taking over from social norms determined by his status as a criminal suspect and a black man in the southern states. This changes the power dynamic in his favour: from low status to high status.

The idea of the conversational exercitive has also been used to explain how slurring utterances are offensive and derogatory (Popa-Wyatt and Wyatt, 2018). When addressing a target with a slur, the speaker's purpose is to grab power by changing the social norms governing the conversation. The mechanism is a conversational

exercitive within a conversational game that assigns a low-power role to the target on the basis of a reference to a low-power historical role held by members of the same group, be that group defined on the basis of gender, nationality, ethnicity, sexuality, religion, disability or another characteristic. This role assignment provides a cognitive shortcut. By using a pre-existing social role and importing it into the conversation, the speaker indexes a suite of oppressive social norms associated with the low-power social role that has been assigned. As we shall see, this role assignment exploits the way that the human brain makes decisions about whether to violate or adhere to a particular norm.

This notion of role assignment is not specific to situations involving power changes. We inherit our roles in discourse from one or more of the many social roles that we each possess in everyday life. If I am a mother and I have a job as a teacher, the norms that come into play when I speak to my child are different to those that are activated when I stand in front of a classroom. We perform a role assignment every time that we introduce ourselves or give someone salient information about our background. We also shape people's views of the social role we fulfill by the way that we interact with them. Finally, and most importantly, the role conveys a great deal of information about the social norms which apply, for remarkably little effort. If I tell you that Jenny is a neurosurgeon with a husband and two children, she lives in California and likes golf, you will have instant access from those five roles to numerous social norms. Whether you attend her clinic, play golf with her, or meet her at a school event, you will have expectations about her behaviour, and possess heuristics to guide yours.

However, there are limitations to the power of the conversational exercitive. If a role assignment carried out by an utterance is a conversational exercitive, then its effects are, by definition, restricted to the conversation. This is because the exercitive act is the illocutionary act and the utterance constitutes the act. Thus, this mechanism cannot, technically, explain effects that persist beyond or occur after the conversation. To illustrate the problem this causes, we consider a modified example of a locker room conversation first noted by McGowan (2004, 2009, 2019). Suppose that the speaker (let's call him Steve) refers to a woman he dated the previous evening (Sue) in a locker-room conversation with his friend (Bob):

(3) **Steve:** '*I banged that bitch last night*.'
Bob: '*She got a sistuh?*'

Now imagine that Bob sees Sue later and engages in a subsequent conversation. Because of the previous conversation, Bob now treats Sue differently than he would have done. This is because his beliefs about her role, and thus as to which social norms apply, have changed. In particular, he will be more inclined to see her as a sexual object and treat her accordingly. So the first conversation had an effect on the second conversation. But the first conversation is over, so there is no score from the first conversation still in existence. So how were the changes propagated? The answer is that Bob carried modified beliefs (be they consciously or unconsciously held) away from the first conversation with Steve. His beliefs will have been determined in part by the role assignment that was made. But these belief changes were not altered automatically by the speech act, since Bob's beliefs are not in the conversational score. Instead, Bob's belief changes are a perlocutionary effect of the role assignment.

A proposal for how those belief changes are made using Bayesian belief updating has been made (Popa-Wyatt, 2024). This proposes that, when a conversational role is assigned, an audience member reasons about hypothesized explanations. One hypothesis is that the role has been assigned incorrectly. Another hypothesis is that the role assignment is correct, and that the target really does possess that social role, from which the conversational role inherits. The Bayesian belief updating rule reasons about the probability of each hypothesis.

So, one way that we can change norms is by re-purposing existing norms to new cases. Role assignment, such as labelling the press the 'enemy' of the American people, is perfect for this. The role is a cognitive shortcut, creating an association between the media and all of the social norms associated with an enemy in the mind of the audience. This begins to answer our second question: 'how does an utterance contribute to altering a social norm that applies in a situation?'. However, it still leaves open the first question: 'why does a speaker make an utterance that (seeks to) alter a social norm?' To answer this, we will need to understand better the game theoretic model of choice between norm adherence and norm violation.

4. Game Theory and Social Norms

As mentioned above, social norms are the collections of informal rules that govern our social behaviour. Social norms are not universal but arise within groups as methods for regulating in-group

behaviour. These norms can be sub-optimal from both an individual and a social perspective. Each of us learns social norms by observing the behaviour of others. Social norms change over time.

There are multiple accounts of how individuals choose whether to adhere to or violate a particular norm in a particular context. A rational choice account is one in which individuals fear fixed social penalties for norm violation and act so as to balance potential penalties and benefits (Coleman, 1994; Axelrod, 1986). This model, however, does not fit with all the available behavioural data. The rational choice model assumes that the decision to adhere to or violate a norm is made in isolation. This is not the case. Instead, norm adherence depends on two kinds of expectations that the individual has (Bicchieri, 2005; Bicchieri and Mercier, 2014). The first type are '*empirical expectations*', i.e., first-order beliefs the individual has about whether others in their group will also adhere to the norm. The second type are '*normative expectations*', i.e., second-order beliefs about whether other group members believe that the individual should also adhere to the norm.

Bicchieri introduced a model of a mixed-motive game that allows a group of players to find a Nash equilibrium balancing these forces (Bicchieri, 2005; Bicchieri and Sontuoso, 2020).[2] In this game, individuals have an expected utility for each possible action (adhere or violate). This expectation is calculated using probabilities that players will adhere to or violate the norm in question. These probabilities are estimated from observations. The game is mixed-motive because the utility combines the material payoff (which typically rises if the norm is violated) and a penalty capturing a psychological cost – or guilt – derived from the maximum cost that another player will incur due to norm violation. Using observations of norm violations to determine subsequent norm adherence yields a better fit to the behavioural data than the rational choice model (Bicchieri and Xiao, 2009). In particular, it fits empirical evidence that shows that there is an asymmetric effect of observed behaviour. In experiments with human subjects, norm adherence declines substantially if the participants observe others violating a norm, whereas norm adherence does not increase substantially if the participants observe other participants adhering to the same norm.

We propose that the mixed-motive aspect – the formulation of the psychological costs of norm violation – can be used to capture why

[2] A Nash equilibrium is a state in the game such that moving out of equilibrium would entail a worse pay-off, so no agent would benefit by changing, given that all other agents don't change (Osborne *et al.*, 2004).

speakers sometimes make utterances that violate social norms. To understand this claim better, let's return to consider the example of Trump's verbal attacks on the media. We'll start with a common-sense explanation of the speaker's motivation and then assess where the game theoretic model requires extension, so as to provide a formal model of that speaker's motivation.

Let's suppose that a speaker wishes to disable a section of the media that they consider unfavourable. Their long-term goals are to sow distrust of, intimidate, cause physical harm to, reduce the audience of, and eliminate scrutiny by that section of the media. In this case, the social norms that the speaker seeks to erode are the social norms of civility, of listening to different viewpoints, and of non-violence in civil society. The speaker also aims to undermine beliefs in media neutrality. Note that the speaker need only erode those norms as applied to the target. This means that norm change by role assignment can be effective. Utterances clearly contribute to this, such as assigning the media the role of an enemy; an enemy who wants to destroy cherished institutions; and an enemy who is lying to achieve their aims. Other utterances would include encouragement to harm individual journalists; to publicly violate norms of politeness when addressing questions; and to verbally threaten those who ask questions. Thus, we can see that, working backwards from the goal of disabling the target, it is rational to make utterances of this nature. The use of role assignment also creates a cognitive shortcut, enabling audience members to use the principle of least effort when making the decision to adhere to or violate the norm (Allport, 1954).

How can this be modelled? What ingredients do we already have? Which ones are still missing? First, let us imagine applying a mixed-motive game directly to the conversation in which an utterance takes place. In this mixed-motive game, the individual would subtract the psychological costs of norm violation from the material benefit. This would require that both the psychological costs and the material benefits can be estimated. Let us focus on estimating the benefit. The speaker needs to be able, first, to define the goal at which they aim. Does that goal state lie within or beyond the conversation?

Our proposal is that speakers aim at a goal for the conversation and place a value on that goal, after having derived that conversational goal from a societal goal. So this, requires that there is a relationship between conversational goals and societal goals. It is not clear how exactly to fill that goal definition gap. There are other gaps. A simple, mixed-motive game is a one-step decision process, with a

single round of play, whereas a dialogue is a multi-step process. An utterance made now will, via the conversational score, affect the dialogue many moves into the future. This is important in a model of norm violation in conversation because I might choose to suffer psychological penalties now (social disapproval) in order to yield a material benefit many steps into the future (disablement of my critics). Yet another gap is that the mixed-motive game is not a model of the conversational dynamics. Nor is it a model of how speakers and listeners generate and reason about sequences of utterances. All it does is provide a way to weigh, across candidate utterances, the pre-calculated long-term benefits and the short-term psychological costs. It doesn't provide a means to estimate the long-term benefits, but merely to employ those estimates.

In this section, we have identified one appealing feature of mixed-motive games for modelling social norm violation: the use of psychological costs estimated based on the observed behaviour of other players. We have also identified several missing elements: multi-step decision making, conversational score updating, and goal definition. We refer to these three gaps as the decision gap, the interpretation gap, and the motivation gap. We now turn to a theory that can provide one of these missing elements. This is a theory of pragmatic inference.

5. Rational Speech Acts

The Rational Speech Act (RSA) framework is a probabilistic – specifically a Bayesian – theory of pragmatic inference. At its core, RSA operates on the principle that speakers are rational agents who aim to be informative, relevant, and efficient in their communication. Listeners, in turn, use these principles to infer the speaker's intended meaning. Starting from a small number of axioms, RSA models both how speakers select utterances and how listeners interpret those utterances. There are significant limitations of RSA. For example, it has been used largely to model the interpretation of single utterances. Nevertheless, it has some utility for our enterprise.

In the RSA framework, the listener maintains a probability distribution over possible interpretations of an utterance. They update this distribution using Bayesian inference that incorporates recursive reasoning to derive the speaker's and the listener's mental models of each other. Specifically, the model incorporates: (i) a model of a literal listener based on the possible semantic interpretations; (ii) a model of a pragmatic speaker that assumes the model of the literal listener; (iii) a

model of a pragmatic listener that assumes the model of the pragmatic speaker. These recursively defined models can incorporate both the costs of an utterance and the prior salience, and thus the probability, of particular interpretations. The significance of RSA is that it fits a variety of human behavioural data for both listeners and speakers.

RSA can, therefore, be used as an ingredient in a model of updates to the conversational score. This is because each participant has a model of the other as a pragmatic listener and so can make updates to the commonly held beliefs. The roles of the participants also sit on the conversational score. Therefore, if a role assignment is made, a psychologically plausible way in which the conversational score updates is to use Bayesian belief updating. Indeed, the Bayesian updating scheme for inferring the social role (which is a belief of the audience member) from the conversational role (which is an element of the conversational score) was proposed by (Popa-Wyatt, 2024), as mentioned earlier.

There are multiple hypotheses or pragmatic interpretations. One is that the target has the social role corresponding to the conversational role. Another is that the target does not. However, determining that variable alone may not be enough to explain the data. To conclude that the target does not have the social role, the listener still requires an explanation of why the speaker made the utterance assigning the corresponding conversational role. Explanations vary according to the context. In the case of the locker room example as in (3), an explanation is that the speaker is bigoted. In the case of Trump's attacks on the media as in (1), it is that he is being insincere so as to gain advantage. The inference could also incorporate Bayesian reasoning to account for inferential bias arising from the degree to which the speaker is trusted by the listener (Asher, Hunter, and Paul, 2021).

This Bayesian updating rule of RSA, applied in an extended way as we propose, can fill the gap of interpretation required to account for how conversational roles are communicated and inferred, but it cannot address the gaps of motivation and decision. We will sketch a further framework for these in the next section.

6. A Sketch of Requirements

We've reviewed three theories. The first was the theory of conversational games and its extension to allow speech acts that update the conversational score so as to change the rules of the conversational game. The second was a game theoretic model of social norm

violation and adherence. The third was Rational Speech Act theory, which is a theory of how utterances are chosen and interpretations of utterances are made. We proposed to use the notion of conversational exercitives from the first; the mixed-motive game with an estimation of psychological costs based on observed behaviour from the second; and the use of Bayesian updating to interpret the meaning of utterances from the third. In the latter two cases, we propose to apply the existing mechanism in a new way, so as to apply them to social norm modelling. We've also identified some remaining gaps, that a complete theory will have to fill. We referred to these as the motivation and decision gaps. Let us give a little more detail on each.

The first remaining gap we termed the motivation gap. This is the problem that, in order to intentionally make utterances that have long-term effects after the end of the conversation, a speaker needs to have a sense of what those long-term effects are intended to be. This requires that the effects are cognitively represented. When a speaker such as Donald Trump attacks the media as in (1), he does so with a clear sense of the long-term disablement it will cause and the value of that to him. But knowing what the long-term goal is, is not enough. It must also be used to derive a goal for the dialogue and a benefit for the current candidate utterance. This is the decision gap.

To solve the decision gap, we require a way for the benefit of achieving long-term social goals to be back-propagated into the current dialogue. One mechanism for this is the theory of stochastic games in which agents play a game, taking decisions in turn, each trying to achieve a long-term goal (Solan and Vieille, 2015). In stochastic games, or multi-stage games generally, participants reason about how to act by attaching rewards to those long-term goals and back-propagating those to estimate the values of actions they can take now. Computational linguists have employed this kind of game-based decision framework to model strategic dialogue planning (Asher and Paul, 2017). These formalisms can provide a framework to explain how long-term motivations are turned into decisions about what to do immediately.

References

Gordon Willard Allport, *The Nature of Prejudice* (Cambridge, MA: Addison-Wesley, 1954).

Nicholas Asher and Soumya Paul, 'Conversation and Games', in *Logic and Its Applications: 7th Indian Conference, ICLA 2017, Proceedings 7* (Springer, 2017), 1–18.

Nicholas Asher, Julie Hunter, and Soumya Paul, 'Bias in Semantic and Discourse Interpretation', *Linguistics and Philosophy*, 45:3 (2021), 393–429.
John Langshaw Austin, *How to Do Things with Words* (Oxford: Oxford University Press, 1975).
Robert Axelrod, 'An Evolutionary Approach to Norms', *American Political Science Review*, 80:4 (1986), 1095–1111.
Cristina Bicchieri, *The Grammar of Society: The Nature and Dynamics of Social Norms* (Cambridge: Cambridge University Press, 2005).
Cristina Bicchieri, *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms* (Oxford: Oxford University Press, 2016).
Cristina Bicchieri and Hugo Mercier, 'Norms and Beliefs: How Change Occurs', in *The Complexity of Social Norms* (Springer, 2014), 37–54.
Cristina Bicchieri and Alessandro Sontuoso, 'Game-Theoretic Accounts of Social Norms: The Role of Normative Expectations', *Handbook of Experimental Game Theory* (Edward Elgar Publishing, 2020), 241–55.
Cristina Bicchieri and Erte Xiao, 'Do the Right Thing: But Only If Others Do So', *Journal of Behavioral Decision Making*, 22:2 (2009), 191–208.
Ken Binmore, *Game Theory and the Social Contract: Playing Fair* (Cambridge, MA: MIT Press, 1994a).
Ken Binmore, *Game Theory and the Social Contract: Just Playing* (Cambridge, MA: MIT Press, 1994b).
James S. Coleman, *Foundations of Social Theory* (Boston: Harvard University Press, 1994).
Noah D. Goodman and Michael C. Frank, 'Pragmatic Language Interpretation as Probabilistic Inference', *Trends in Cognitive Sciences*, 20:11 (2016), 818–29.
David Lewis, 'Scorekeeping in a Language Game', *Philosophical Papers*, 1 (1979), 233–49.
Aisha Majid, 'Trump vs media: Four years of presidential press attacks charted', in *Press Gazette*, Future Of Media (2022).
Mary Kate McGowan, 'Conversational Exercitives: Something Else We Do With Our Words', *Linguistics and Philosophy*, 27 (2004), 93–111.
Mary Kate McGowan, 'Oppressive Speech', *Australasian Journal of Philosophy*, 87:3 (2009), 389–407.
Mary Kate McGowan, *Just Words: On Speech and Hidden Harm* (Oxford: Oxford University Press, 2019).

Cailin O'Connor, *The Origins of Unfairness: Social Categories and Cultural Evolution* (Oxford and New York: Oxford University Press, 2019).
Martin J. Osborne, *An Introduction to Game Theory* (New York: Oxford University Press, 2004).
Mihaela Popa-Wyatt, 'Norm-Shifting through Oppressive Acts', in Sally Haslanger, Karen Jones, Greg Restall François Schroeter, and Laura Schroeter (eds), *Mind, Language, and Social Hierarchy: Constructing a Shared Social World* (Oxford: Oxford University Press, 2024).
Mihaela Popa-Wyatt and Jeremy L. Wyatt, 'Slurs, Roles and Power', *Philosophical Studies*, 175:11 (2018), 2879–2906.
Brian Skyrms, 'Signals, Evolution and the Explanatory Power of Transient Information', *Philosophy of Science*, 69:3 (2002), 407–28.
Eilon Solan and Nicolas Vieille, 'Stochastic Games', *Proceedings of the National Academy of Sciences*, 112:45 (2015), 13743–46.
Kim Sterelny, *The Pleistocene Social Contract: Culture and Cooperation in Human Evolution* (Oxford: Oxford University Press, 2021).
Lynne Tirrell, 'Genocidal Language Games', in Mary Kate McGowan and Maitra Ishani (eds), *Speech and Harm: Controversies over Free Speech* (Oxford: Oxford University Press, 2012), 174–221.