Opinion

CellPress

Unwritten rules: virtual bargaining underpins social interaction, culture, and society

Jennifer B. Misyak[1], Tigran Melkonyan[1], Hossam Zeitoun[2], and Nick Chater[1]

[1] Behavioural Science Group, Warwick Business School, University of Warwick, Coventry, UK
[2] Strategy and International Business, Warwick Business School, University of Warwick, Coventry, UK

Many social interactions require humans to coordinate their behavior across a range of scales. However, aspects of intentional coordination remain puzzling from within several approaches in cognitive science. Sketching a new perspective, we propose that the complex behavioral patterns – or 'unwritten rules' – governing such coordination emerge from an ongoing process of 'virtual bargaining'. Social participants behave on the basis of what they would agree to do if they were explicitly to bargain, provided the agreement that would arise from such discussion is commonly known. Although intuitively simple, this interpretation has implications for understanding a broad spectrum of social, economic, and cultural phenomena (including joint action, team reasoning, communication, and language) that, we argue, depend fundamentally on the virtual bargains themselves.

Introduction

Unlike many animal species, humans do not face an individual battle against the natural environment. Instead, we have collectively constructed an enormously intricate social environment requiring complex coordination of behavior between individuals, within organizations, and throughout entire societies. According to perspectives from across the social sciences, as well as folk intuition, the coordination of behavior between individuals and institutions is governed by astonishingly rich patterns. Such patterns for coordination include: norms, customs, and conventions governing the interpretation of language, gesture, and facial expressions; 'scripts' for any number of standardized interactions (e.g., being seated and ordering at a restaurant); social and organizational roles; organizational structures (e.g., hierarchical management); and shared esthetic and moral evaluations. Such patterns can also arise extemporaneously and flexibly (e.g., in the joint action of moving a sofa, I pull while you push; in language understanding, we both take 'Leonardo' to refer presently to da Vinci, but moments later to an uncle of the same name).

Corresponding author: Chater, N. (nick.chater@wbs.ac.uk).
Keywords: virtual bargaining; social interaction; culture; joint action; human communication; coordination.

1364-6613/
 http://dx.doi.org/10.1016/j.tics.2014.05.010

What is the nature of these unwritten rules that coordinate human behavior? What is their origin? How are they learned? We propose that these rules can be viewed as resulting from virtual bargains; that is, from agreements that social participants anticipate they would make, were they to engage in explicit bargaining. However, although the results that would arise from such discussion are 'obvious' to participants (although a theory will be required, of course, to explain what is obvious), actual discussion is unnecessary; participants can each act as if such discussion occurred and coordinate their behavior appropriately. Complex social patterns arise because past virtual bargains can help shape similar future bargains, and bargains can layer on, and interact with, each other.

The virtual-bargaining account operates within the framework of rational-choice theory, which is prevalent in economics and social science and employed in 'rational' models of cognition. In the cognitive sciences, many models explore restrictions on rationality (e.g., due to computational limitations). Here, by contrast, rather than focusing on bounded rationality, we take a complementary approach: we extend the scope of rational-choice models of interaction. Virtual bargaining provides a bridge between individual cognition and the unwritten rules that underpin social interaction, culture, and society.

Let us start by considering a social world without unwritten rules as basic elements, asking how people might coordinate their behavior and how unwritten rules might arise from such coordination. Along the way, we will consider implications for social behavior and sketch how the elaborated layering of these unwritten rules can generate culture and social and political structures.

How can individualistic agents coordinate?

Consider perhaps the simplest possible coordination problem, the Hi–Lo game (Figure 1) [1], in which two players must simultaneously and independently choose between 'Hi' or 'Lo'. Only if both choose identically do they receive any pay-off, which may be either large (both choose Hi) or small (both choose Lo). It is intuitively obvious, and borne out by experiment [1], that both players will choose Hi, but explaining even this trivial case of coordination is difficult for many models in cognitive science. In such symmetrical

Glossary

Common knowledge in philosophy and rational-choice theory, X is said to be common knowledge between agents A and B if they both know X, each knows that the other knows X, each knows that the other knows that they know X, and so on indefinitely. Less-stringent definitions are frequently used in discourse analysis, psychology, and other fields (for an overview, see [20]); the key aim is designating the set of knowledge, beliefs, assumptions, and/or experiences that agents A and B mutually recognize as shared and that are drawn on for real-time understanding in social interactions.

Focal points in a game with many Nash equilibria, one or more of these equilibria may appear particularly salient, or 'focal'. For example, if two people must independently choose the same letter of the alphabet without communication, they are much more likely to select 'A' than 'K'. A challenge for future research is to uncover the cognitive and cultural bases of focality.

Joint action disciplines vary in their definition of joint actions – from philosophical accounts that invoke intentionality, to psychological perspectives that are broader in their usage and, for example, may focus on the behaviors of coupled systems (for overviews of special journal issues on the topic, see [21,22]). According to Bratman [23], joint action arises when two or more people act according to a shared plan that coordinates their behavior. Virtual bargaining provides a possible mechanism by which the plan can be 'agreed on' without explicit communication.

Joint attention/perception we jointly attend to X if the fact that we are attending to X is common knowledge between us. That is, not only are we both, say, looking at the same object, but we both know that we are looking at the same object, and know that we know this, and so on.

***k*-rationalizability/rationalizability** these generalizations of Nash equilibrium maintain the assumption of common knowledge of rationality but relax or abandon the assumption of rational expectations [24,25]. *k*-rationalizability reflects finite levels of iterated knowledge of rationality. A 1-rationalizable strategy is a best response to some profile of other players' strategies whereas a *k*-rationalizable strategy is a best response to some $(k-1)$-rationalizable strategy profile of other players. An infinitely repeated iterative process ($k=\infty$) corresponds to rationalizability and reflects common knowledge of rationality with no further restrictions on beliefs. In other words, a strategy is rationalizable if a perfectly rational player could justifiably play it against perfectly rational opponents.

Level-*k* models these are in many respects similar to rationalizability. In level-*k* models [26–28], a level-0 player represents a non-strategic type that follows some exogenously prescribed behavior, a level-1 player best responds to level 0, and so on. Thus, players' types are rational in the sense of best responding to some beliefs; they depart from equilibrium only in that the beliefs are based on the simple, nonequilibrium models of others. A level-*k* player can perform at most *k* 'steps of reasoning' in the sense that she can iterate the best response correspondence at most *k* times. Thus, in this sense, level-*k* players respect *k*-rationalizability. However, *k* can vary across players in level-*k* models and, consequently, the latter can be viewed as a heterogeneity-tolerant refinement of *k*-rationalizability.

Nash equilibrium the canonical model of strategic reasoning, defined as a profile of strategies, one for each player, such that each player's strategy maximizes her expected pay-off, given that the other players follow their parts of the equilibrium prescription. Nash equilibrium is a considerably stronger requirement than rationality in the usual decision-theoretic sense. It also entails a rational-expectations assumption that all of the players' beliefs about each others' strategies are correct. Systematic deviations from Nash equilibrium predictions in laboratory, field, and naturally occurring settings have led to the development of several alternative theories of strategic reasoning; the most notable among these are rationalizability, quantal-response equilibrium, and *k*-level reasoning/cognitive hierarchy.

Quantal-response equilibrium a model of strategic reasoning in which players are assumed to make errors in choosing which pure strategy to play. The probability of any particular strategy is increasing in its expected pay-off (so that costlier errors are less likely) and is evaluated taking the noisiness of other players' strategies into account [29].

Team reasoning a form of reasoning in which agents infer, and implement, a shared plan that is 'best for the team' (e.g., [30,31]). Team reasoning, like the related notion of 'we-thinking' [32,33], is often viewed as a particular 'frame of mind': people can choose to be individualists or team reasoners, depending on the context. Virtual bargaining is a possible account of the origin of team 'objectives' and 'plans', but virtual bargaining is a broader notion, applying even when people have conflicting interests (as might be true of the despot and servant in Box 3).

coordination games, each player should follow whatever their coplayer does.

There is, however, a danger of infinite regress. Suppose, for example, we apply a simulation theory of mind [2,3]. Player 1 attempts to ascertain what Player 2 will do by using Player 1's own cognitive machinery to simulate Player 2. However, the 'simulated' Player 2 now has to simulate what Player 1 will do – and this is precisely what Player 1 was trying to determine by performing the simulation of Player 2 in the first place. A theory theory of mind [4,5] fares no better. Player 1 cannot determine her own move using an intuitive theory of Player 2's mind because any such theory must presumably recognize, by symmetry, that Player 2 is attempting to apply an intuitive theory to Player 1, and so on indefinitely.

	Player 2: 'Hi'	Player 2: 'Lo'
Player 1: 'Hi'	10, 10	0, 0
Player 1: 'Lo'	0, 0	1, 1

TRENDS in Cognitive Sciences

Figure 1. The Hi–Lo game. The players receive nothing if they choose different moves. Depending on which move they 'agree on', they each receive either $10 or $1. In these and other games, the labels 'Hi' and 'Lo' are used to assist the reader; they are not presented to experimental subjects.

Because each player's optimal actions are inextricably dependent on the (unknown) actions of the other, this 'loop' of circular reasoning cannot be broken even if we recognize players to have 'social' or other-regarding preferences [6,7], such as altruism, that reflect a concern for the coplayer's outcomes. Crucially, the problem of infinite regress is very general and arises for a wide spectrum of cases wherein people must coordinate their behavior to mutual advantage, including, we suggest, many interesting aspects of social behavior, such as joint action and communication, that lay the foundations for human culture.

Coordination, social equilibria, and virtual bargaining

One way to break out of the infinite regress is to switch focus from the cognitive operations of each agent to the possible equilibria of the 'social system' – here, the moves of the pair of agents. Nash [8] considered an equilibrium to be a pair of moves (or probability distributions over the moves of each agent) that arises when neither agent has incentive to switch moves in the light of the other's equilibrium move.

Coordination problems distinctively have multiple Nash equilibria; including here, (Hi, Hi) and (Lo, Lo). However, the equilibrium (Hi, Hi) is preferable to (Lo, Lo) for both players. This suggests a strategy for both agents to coordinate successfully: where it is obvious which Nash equilibrium is 'mutually agreeable', in some sense both players should choose their moves to achieve that equilibrium.

So how do players know on which equilibrium to converge? Broadly, we propose that people coordinate by virtual bargaining [9]; that is, by imagining which equilibrium would be chosen if the players could communicate and explicitly make offers and counter-offers until a bargain is reached. In the Hi–Lo game, it is clear that any reasonable process of bargaining would favor (Hi, Hi) over (Lo, Lo).

The idea of virtual bargaining can be made formally precise in numerous ways, but in relation to pay-off structures virtual bargaining will entail identifying the set of

Box 1. Defining a virtual-bargaining equilibrium

Virtual bargaining involves a psychological claim: that some, and perhaps many, social interactions involve people behaving as if they had been able to bargain about what to do. Aspects of virtual bargaining can potentially be formally modeled in various ways. Here we explore one possible approach, using game theory.

To formally define a virtual-bargaining equilibrium, consider a two-player game with simultaneous and independent strategy choices by the players. A feasible agreement is a strategy profile such that neither player's most beneficial deviation from the agreement (if any) is at the expense of her opponent. When making their choices, the two players simulate a bargaining process that, given the players' *status quo* positions, chooses one of these feasible agreements. The set of feasible agreements includes all Nash equilibria. The *status quo* position is one of (possibly many) Nash equilibria (although due to space constraints, we do not detail the specifics of selecting a *status quo* Nash equilibrium when there are multiple Nash equilibria) and is the fallback position if no agreement is reached. The bargaining mechanism used to choose a virtual bargain is the Nash bargaining solution (although one can allow for alternative bargaining mechanisms to arrive at the virtual bargain): this is the pair of strategies that maximizes the product of differences between the players' expected utility under this strategy pair and the expected utility from the *status quo*.

The psychological notion of virtual bargaining is consistent with Schelling's explanation that many situations 'provide some clue for coordinating behavior, some focal point for each person's expectations of what the other expects him to be expected to do' ([34], see p. 57). However, the technical notions of virtual bargaining and focal-point reasoning diverge. Focal-point reasoning typically considers coordination by using labels for different options rather than pay-offs (e.g., a choice of a uniquely colored red might be preferred). Labels play no role in our formal definition of virtual bargaining, which focuses purely on pay-off structure. Even when focal-point reasoning stems from the pay-off structure, existing theoretical accounts (e.g., the principle of collective rationality in [35]) differ from virtual bargaining.

Empirically, focal-point reasoning applies best to games where conflicts of interest are small ([35], see p. 548) and have limited explanatory power under pay-off asymmetries [36]. Identifying conditions on pay-offs (e.g., in terms of common versus conflicting interests) under which virtual bargaining has the most empirical bite remains an open challenge.

'feasible agreements' [9] (Box 1). As we delimit these here, such feasible agreements preclude opportunities for 'exploitation'; that is, a bargain will not be 'feasible' if one player's most beneficial deviation from the 'agreed' move would result in unilateral gains for the switching player and losses for the coplayer. So, in the classic game of Prisoner's Dilemma, a virtual bargain where we both agree to cooperate may be undermined because either player can exploit the other by defecting, and hence benefiting, at the other's expense. The feasibility criterion makes few assumptions about the structure of the 'game' (i.e., no need for a third party to enforce virtual bargains) or about relations between players (i.e., no requirements for trust or other-regarding preferences). The set of feasible agreements will typically encompass a broader range of possible agreements that includes standard Nash equilibria (e.g., [9]).

Thus, in explaining how individuals coordinate their behavior, the rudiments of a formal account of virtual bargaining have the following distinct components: (i) the notion of a feasible agreement broadens the set of equilibria beyond Nash equilibria; and (ii) the virtual bargaining equilibrium is a selection from among the feasible agreements. Where it is obvious to both players that a particular feasible agreement would be chosen, the players can choose it directly and circumvent the problem of infinite regress. Notably, virtual bargaining may enable people to coordinate even when the players are antagonistic to each other, with conflicting rather than perfectly coinciding interests. We note, however, that although pay-off structures are one canonical way of modeling interdependent interactions among individuals, future research may also extend the formal theory of virtual bargaining to wider classes of agreements (e.g., those in which decision making depends on decision labels, not pay-offs).

Many human interactions are almost as straightforward as the Hi–Lo game. In deciding who should push and who should pull a table across a room, one arrangement may be obviously better for each person (in terms of, say, walking distance to take up their positions, as in Figure 2A). Figure 2B shows a scenario whereby both people are closer to the left-hand side of a table, but the most natural way each moves to take hold of the table minimizes the joint distance moved. Any number of other factors can also affect the virtual bargain that people will often effortlessly and instantaneously strike. If one person moves more slowly or one end requires more force, people can often shift appropriately. The coordination outcome is clear to both parties and no communication is required to coordinate successfully.

From this account, virtual bargaining mirrors actual bargaining and the multitude of factors determining such bargains. Differences in power or social status provide one particularly prevalent source. So, for instance, in the unlikely event that a despot and servant were moving a table together, as in Figure 2C, the despot (who we assume here to be powerful, selfish, and lazy) has no need to instruct the servant to take the longer route. Note that even when power is so sharply asymmetrical, virtual bargaining is required (because the despot also needs to know what the servant will do).

Specifically, the factor of 'bargaining power', as in explicit bargaining, influences the virtual bargain that is reached and depends on the pay-offs of the game. In our example, the despot has more bargaining power, presumably due to the ability to exact terrible consequences on a disobedient servant. Another important factor for virtual bargains involves other-regarding preferences, which may be incorporated by modifying the game pay-offs to reflect concern for the other. If the despot and servant become friends, it might become common knowledge that both parties wish to find a 'bargain' that is equally good for both. Given its game-theoretic foundation, the theory of virtual bargaining can capture these relevant factors using established game-theoretic models of explicit bargaining.

The origin of unwritten rules

In this perspective, the starting point for the unwritten rules governing social interactions is the virtual bargains themselves. Such virtual bargains (like customs, conventions and, social norms) have normative force: as with an explicit bargain (e.g., as codified in a legal agreement), participants know that they should behave as the bargain specifies. (Although, as with real bargains, people can

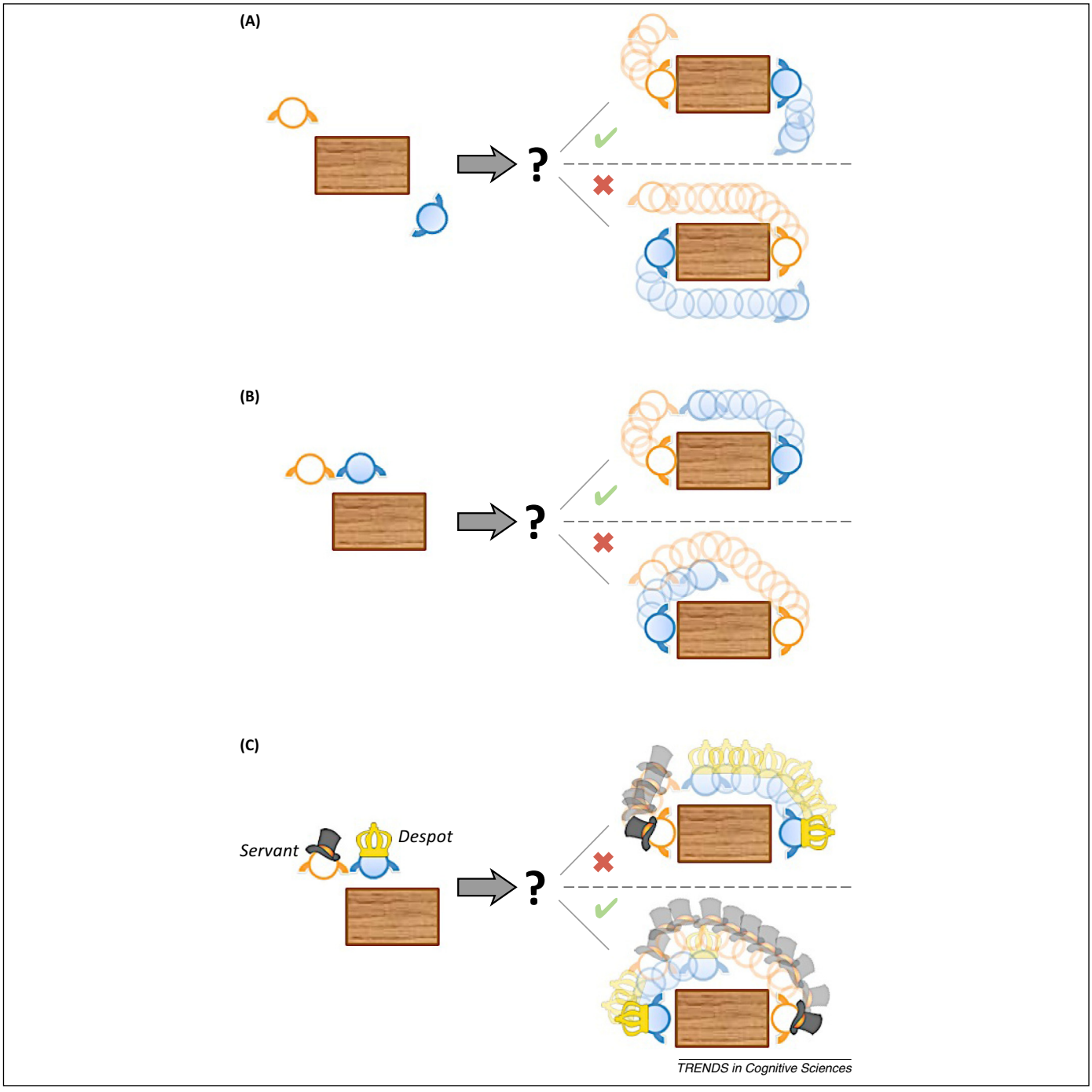

Figure 2. Two people (unfilled orange and shaded blue circles) have to coordinate on moving a table from various initial locations. In both **(A)** and **(B)**, the natural configuration seems to be for the orange circle to go to the left and the blue circle to the right (as depicted in the top, rather than bottom, scenarios). In (A), analogous to the Hi–Lo game, this option is preferable for each person individually; in (B), both persons are nearer to the left-hand side of the table, but the top configuration minimizes joint distance moved. However, other factors may affect which bargain would be reached and hence which virtual bargain the individuals infer. In **(C)**, for instance, a 'despot' and 'servant' both know that any 'bargain' between them would simply perform the despot's wishes. If the despot is mutually known to be 'lazy' (i.e., to wish to minimize her own effort), both parties can immediately reach a virtual bargain opposite to that in (B) without communication.

recognize the appropriate virtual-bargaining solutions without the intention of following them.)

What is the content of a virtual bargain? A natural, albeit partially adequate, answer is that the contents correspond to an 'explicit agreement' to which both parties are committed. These contents may be complex (e.g., 'I'll take this end of the table, you'll take that end, and then we'll both lift it, and then I'll shuffle backwards and you forwards'), with endless clauses (e.g., 'on encountering stairs, I'll lift my end of the table and back up the stairs, one by one', 'neither of us will intentionally drop the table without warning', 'we undertake to continue until the table is at its new destination').

Such bargains can never be made fully explicit because any number of possible eventualities must be catered for. Indeed, written contracts are inevitably open ended, or 'incomplete' (in legal and economic terminology), due to an infinity of unforeseen contingencies (e.g., both parties are

permitted to cease moving the table if it threatens to break, there is a fire alarm, or the way ahead is blocked by an unanticipated cocktail party). Clearly, not all such eventualities can be anticipated: any bargain may require further (virtual or real) bargaining at some later stage. Where virtual bargaining is not possible (i.e., what 'we' would agree to do is nonobvious to both of us), people will often shift to real bargaining; that is, explicit discussion of next actions. Communication is required just in so far as bargaining can allow joint behavior to proceed (e.g., a nod or glance might suffice to indicate that we twist the table clockwise rather than counterclockwise before ascending steps).

In this way, it is instructive to note that virtual bargaining also addresses a foundational issue in pragmatics: the rationale by which, given their common knowledge, speaker and hearer can coordinate on one of many possible messages that a given highly ambiguous signal might provide (Box 2). The core idea is that speaker and hearer can 'virtually agree' many aspects of the use of a communicative signal in the light of their common knowledge (for example, that 'Could you tell me the way to the station?' is almost always indicative of a desire to go to the station and hence requires more than a yes/no answer). A natural challenge for future work, therefore, will be to explain how traditional pragmatic notions such as 'quantity', 'quality', and 'relevance' [10–12] may be reconstructed within the virtual-bargaining framework.

Box 2. Communication, language, Schelling, and virtual bargaining

Nobel prize-winning economist Thomas Schelling was one of the first to explore how well people are able to coordinate their behavior, given a minimal clue [34]. He gave people the hypothetical problem of meeting with somebody in New York City on a set date, but with no time or location specified and no opportunity to communicate. Participants converged on a small number of 'focal' solutions, the most popular being midday at Grand Central Station. In terms of the present analysis, participants are attempting to infer a virtual bargain: what would be a sensible time and place to agree to meet.

It is a small step to recognize that the same situation arises when the 'clue' that must be solved (to meet tomorrow in New York City) is given not by a third party but by one participant to the other: 'Let's meet tomorrow in New York'. In the absence of further interaction, the participants will, as before, coordinate successfully when they infer the same virtual bargain. Thus the process of interpreting language, already a communicative signal, involves virtual bargaining between speaker and hearer.

Felicitous communication typically requires that the communicative clue is sufficient that the virtual bargain uniquely determines what is to be conveyed. The clue of Schelling's example is unlikely to be sufficient for a reliable meeting to be coordinated, but if, for instance, the speaker and hearer are members of the same company with London and New York offices and a history of prior lunchtime meetings, it might suffice.

Virtual bargaining provides a novel starting point for providing foundations for Grice's program (e.g., [10]) explaining how minimal communicative and linguistic signals can, by sophisticated inferential processes, convey remarkably rich messages. From the present perspective, both speaker and hearer face the same inferential problem: to infer what both would virtually agree, given that signal. Virtual bargaining provides a framework for integrating and making precise the interplay of specific pragmatic principles (e.g., principles of cooperative and relevance, maxims of quality and quantity; for related approaches, see [11,12]; see also [37]).

So we have, then, a partial answer to the origin of the unwritten rules of social interactions. Such rules need not be prespecified by some external authority and learned by members of a group or culture. They can be composed as we go along, by engaging in ongoing virtual bargaining. Furthermore, these unwritten rules are complex due to the open-ended nature of our ability to virtually agree what to do in different situations.

Each new case of virtual bargaining does not stand alone, but depends on prior social interactions in at least three important ways.

First, a history of social interactions, communication, and common culture can foster common knowledge among individuals, thereby facilitating virtual bargaining (Box 3). Virtual bargaining will go astray if, for instance, partners disagree about where they are aiming to put the table or which person is stronger/which end is heavier. Moreover, mere mutual knowledge (i.e., both players knowing the same thing) is insufficient; for each player to be able to justify following the terms of a virtual bargain, the players must have common knowledge of the assumptions on which this bargain is based; that is, each knows these assumptions, knows that the other knows, knows that the other knows that they know, and so on indefinitely (e.g., [13,14]). For example, if both players believe that the table is heading for the living room but one erroneously believes that the other has a different goal, virtual bargaining is likely to fail.

Second, previous virtual bargains (and, more broadly, real bargains and past outcomes) provide precedents for current and future virtual bargains, just as previous legal judgments provide precedents for today's court cases. If the stronger partner lifts the heavy end of the table, both parties are likely to take this precedent into account in future bargains, perhaps generalizing to other cases, such as who carries the large versus the small suitcase. Similarly, if the singer and support band split the money 50–50 at the last gig, both parties are likely to assume the same division applies for future gigs – and either party is likely to be outraged if the other demands, after the fact, more favorable terms. Thus, past (virtual) bargains may rapidly become entrenched and generalized.

Coordination is likely to follow precedents for two reasons. One relies on virtual bargaining: common knowledge of the precedent marks that coordination solution as 'special' and thus acts as a possible tiebreaker for choosing between future solutions. Here the logic is that precedence is acting the same way as any other salience marker of which we have common knowledge. So, for example, suppose that two people have to coordinate on choosing the same digit written on rectangular cards; having coordinated successfully on, say, '6' previously makes this choice salient, in the same way as if that digit alone were presented in bold font or presented on a triangular card. However, precedents may be followed even if one or both participants do not continue to apply the logic of virtual bargaining; people may reflexively choose the precedent merely because it is salient, irrespective of whether this is common knowledge. These possibilities can be distinguished empirically by varying relevant common knowledge. For example, 'mere salience' should drive choice

Box 3. Virtual bargaining, joint action mechanisms, and team reasoning

Virtual bargaining depends on common knowledge. Suppose, for instance, that the servant and the despot are moving a table, but only one of them knows that the end of the table nearest the despot is sticky with fresh paint. If this were common knowledge, the servant would choose the sticky end, and the despot would choose the other end. However, given their current common knowledge (which does not include the crucial information about the sticky paint), the opposite virtual bargain is likely.

The problem is most readily resolved by one party bringing the crucial information into common knowledge. For example, if the despot has the crucial information, she might pointedly glance or gesture at the sticky paint, with the expectation that the servant will follow that look or gesture and that they will jointly attend the sticky paint [38–40]. Aspects of the external world that are jointly attended enter into common knowledge. The virtual bargain will then be revised and the servant will rush dutifully to the sticky end. However, the power asymmetry allows that the despot can instead simply begin to move toward her preferred end and, whatever the servant's initial plans, the servant will follow her lead. If, by contrast, the servant, but not the despot, has noticed the sticky paint, a glance or gesture may not suffice; establishing common knowledge will probably require direct communication (say, 'Madam, I believe the paint at this end is not completely dry').

The manipulation of joint attention is itself a critical element in social interaction: joint attention influences common knowledge, which in turn may influence the virtual bargain reached. Thus, joint attention is a resource that must be virtually bargained over and is subject to subtle cultural conventions; for X to draw Y's intentionally directed attention to something requires joint agreement and is governed by conventions such as turn taking, eye contact and ostensive eye gaze, and head movements, as well as being strongly determined by power/status asymmetries [41–46].

Whereas joint attention may help establish common knowledge required to facilitate virtual bargaining, joint action can result from virtual bargaining. From the present perspective, a joint action can be defined simply as the implementation of the same virtual bargain by two or more participants (this viewpoint is closely related to [23,47]). In particular, virtual bargaining can be viewed as a type of team-reasoning account of social interaction [47], where the virtual bargaining process specifies the actions of the team, based on the preferences and beliefs of the individuals in that team. Note, however, that virtual bargaining can occur between friends or adversaries, whereas team reasoning, as the name suggests, is often taken to apply to team members with overlapping interests.

Philosophical approaches typically define joint actions with respect to intentionality (whether shared or collective; e.g., [48–50]), whereas psychology and neuroscience often favor broader terms, by which joint actions are understood more simply as the result of cognitive, perceptual, and motor mechanisms that facilitate emergent coordinative structures (e.g., [51,52]). Our definition thus follows the 'tighter' usage of theoretical approaches and is sharply distinct from spontaneously organized behavior and from coordination mechanisms, which operate by, for example, priming one person's motor [17], language [17], or emotional [53] system by the observation of another. Nonetheless, joint actions and coordinative mechanisms may interact in interesting ways.

when Person 2 attempts to coordinate with Person 3, even if the precedent was set when Person 2 interacted with Person 1 (and is thereby not common knowledge for Persons 2 and 3). Conversely, if it is common knowledge (of a prior coordinative solution) that supports following a given precedent, virtual bargaining should be sensitive to direct changes to common knowledge, in contrast to mechanistic approaches. In such a case, as another example, if it is known that one partner suffers from relevant memory loss or impairments, the other partner is unlikely to use a precedent (on the merits of its shared historical status) as an appropriate tiebreaker for future bargains.

Virtual bargaining may therefore comprise part of the answer to where norms come from and how conventions are established, although fleshing out that account may be complex. Drawing on precedents when a new situation is identical with a previous situation is relatively straightforward. In general, however, the new situation will link with numerous, possibly conflicting precedents; virtual bargaining may be needed at a higher level and possibly require substantial reasoning and generalization, analogous to that in a precedent-based legal system, for individuals to converge on choosing the 'right' precedent.

Third, and last, virtual (and real) bargains can be layered on top of one another through processes of social interaction and cultural evolution, just as formal institutions such as banks, courts, or police forces are defined by layers of explicit legal agreements. For example, banking requires legal definitions of debtors' and creditors' responsibilities, the existence of money with legal agreements concerning the operation of a currency by a central authority, legal definitions of property rights, and the definition of the individual as a legal entity. Similarly, informal 'institutions' (such as friendship, membership within religious or cultural groups, or even sports and games) may require layered virtual bargains of substantial complexity. For instance, playground soccer (with no official rule book) requires children to agree what the goal is (often defined, for example, by pre-existing marks on the ground or articles of clothing; lacking a crossbar, the height of the goal can sometimes be disputed), what counts as scoring a goal, when the ball is 'in play', the game's objective (e.g., first to five goals), who the teams are (including designated goalkeepers), and more basic agreements concerning what it means to be in a team. These systems of unwritten rules provide the basis for momentary agreements that may be required for joint actions such as one player passing the ball to another (in contrast to one player kicking a ball and another happening to intercept it).

Virtual bargaining thus provides an account of the nature of the implicit rules that govern social interaction. It also provides a starting point for thinking about how such rules, the cultural patterns they support (e.g., conventions of politeness [15]), and social, organizational, or political structures (e.g., family units, hunter–gatherer bands, sports teams, clubs, companies, governments, countries) may emerge – through incremental layering of successive bargains. Box 4 explores these connections further.

Future directions

A key challenge for future work is building a formal theory of virtual bargaining. One approach is to divide the task into two principal components (as outlined in 'Coordination, social equilibria, and virtual bargaining' above): an account of which virtual bargains are, in some sense, self-reinforcing (given that there is no third party to enforce virtual bargains); and a theory of the virtual-bargaining process, to determine which virtual bargain might be agreed on. Misyak and Chater [9] sketch a possible approach (corresponding with slight modification to the

Box 4. From virtual bargains to culture and society

We suggest that the ability to create virtual bargains underpins the complexity of society and that the open-ended, flexible, and contingent nature of such bargaining helps explain how we are able, to a good approximation, to assimilate and follow what might appear to be an inordinately elaborate set of rules without consulting any explicit rule book.

Yet, of course, such rule books exist: games and sports have official rules and organizations of all sizes have more-or-less-formal codes of conduct, articles of association, and written constitutions. We have already suggested that unwritten rules can never be made fully explicit, although a truly staggering volume of laws and regulations has been formulated and adopted in industrialized societies [e.g., in the USA alone, the 2012 annual edition of the Code of Federal Regulations contained 174,545 pages spanning 235 volumes (US Office of the Federal Register, *Federal Register & CFR Publication Statistics*)]. In many cases, the explicit formalization of rules of behavior follows the well-trammeled lines of previous practice, so that 'formal' agreement is typically patterned on the implicit rules immanent in existing virtual bargains.

Note, however, that the process of formalization need not necessarily codify existing practice but may modify it, sometimes substantially. For example, when the US National Hay Association standardized trade rules, it purposely deviated from some of the customary practices to promote uniformity and what it considered 'best practices' [54]. Similarly, the formulation of prescriptive grammar can modify the unwritten rules that guide speakers of a language, although such modifications are typically fairly marginal (e.g., impositions to avoid splitting infinitives or 'stranding' prepositions).

Notice, too, that the accretion of rules is not merely the passive layering of one cultural 'sediment' over another. Instead, it is often a highly interactive process, such that later 'bargains' may modify interpretation of earlier bargains (e.g., the notion of property has required continual revision to deal with money, debt, intellectual property, and share ownership). Moreover, new virtual bargains may abolish or eliminate previous bargains. For instance, new norms about the equality and treatment of minorities may clash with and partially suppress or improve unwritten rules within the dominant culture.

It is tempting to suppose that, in a world of legislatures, lawyers, and regulators, virtual bargains play a decreasing role in society; might virtual bargains be, to adapt Wittgenstein's phrase from another context, a ladder that we can kick away once society has sufficient apparatus for formulating rules explicitly? We suggest that this is highly misleading for three reasons: (i) rules, even where available, are, in many contexts, rarely consulted – for example, few academics, we suspect, have ever scrutinized their terms of employment, their job description, or the constitution of their University; (ii) explicit rules are invariably incomplete and can be applied to new circumstances only by the application of virtual bargaining, which is especially important in the context of rapid cultural and technological change; and (iii) communication – and specifically language, in terms of which explicit rules are formulated – presupposes virtual bargaining (Box 2).

formal definition in Box 1) which is based on game theory, using Nash's theory of bargaining, but broadening the notion of a Nash equilibrium as a self-reinforcing equilibrium. However, virtual bargaining can be developed in a wide variety of ways.

A further challenge is empirical: what are the key experimental tests for distinguishing the virtual bargaining account of how people coordinate their behavior from related approaches, such as team reasoning (e.g., [1]; see also Box 3 and Glossary), and game-theoretic approaches such as quantal-response equilibria, level-*k* reasoning, and rationalizability? In addition, which aspects of joint action, communication, and social interaction are best explained by virtual bargaining rather than by more mechanistic accounts of behavioral coordination (e.g., [16–18])?

If social interactions are governed by virtual bargains, we might anticipate that many of our moral intuitions, and perhaps many of our social emotions, will be defined not in terms of outcomes but in terms of adherence to, or violation of, virtual bargains. This might be one source of the strong moral and emotional feelings of young children in relation to the sharing of food or toys, the emphasis of trustworthiness as a central moral virtue, and the intuitive basis for deontological (rule-based) ethical theories (e.g., [19]).

Thus, many new questions emerge for understanding and developing the theory of virtual bargaining (Box 5). If, as we have argued, virtual bargaining underpins the unwritten rules of society, the virtual-bargaining perspective should in turn promise new insights toward elucidating a correspondingly broad range of human cultural and societal phenomena. In particular, we have noted some starting points for considering implications in ethics, psychology, economics, communication, and even language, among others, with the virtual bargains themselves suggesting a fresh source of fruitful interconnections across these literatures.

Box 5. Outstanding questions

- Which social interactions require virtual bargaining and when can low-level 'proxies' suffice?
- How does virtual bargaining relate to the problem of generalizing from past experience?
- How far does allowing players to communicate before making their decisions (so-called 'cheap talk' [55]) relate to the results of virtual bargaining?
- How may common processing factors and shared sociocognitive contexts influence virtual-bargaining outcomes?
- How can the formal theory of virtual bargaining be extended to deal with wider classes of agreement? For example, how can a formal theory of focal-point reasoning, which depends on decision labels and not pay-offs, be constructed in this framework?
- How far do past virtual bargains become embedded in future interactions?
- How do we establish the common knowledge required for virtual bargaining?
- How does virtual bargaining help us understand how people attribute responsibility to individuals, when outcomes depend on the actions of many people?
- Which aspects of moral thinking are rooted in virtual bargains and their violation?
- What is the developmental trajectory of virtual bargaining?
- Is virtual bargaining uniquely human?

Acknowledgments

J.B.M. and N.C. are supported by European Research Council (ERC) grant 295917-RATIONALITY; N.C. is partially supported by the Economic and Social Research Council (ESRC) Network for Integrated Behavioural Science, the Leverhulme Trust, Research Councils UK Grant EP/K039830/1, and the Templeton Foundation. The authors thank Max Kleiman-Weiner, Josh Tenenbaum, and an anonymous reviewer for valuable feedback on an earlier version of the article.

References

1 Bacharach, M. (2006) *Beyond Individual Choice: Teams and Frames in Game Theory* (Gold, N. and Sugden, R., eds), Princeton University Press

2 Gordon, R. (1986) Folk psychology as simulation. *Mind Lang.* 1, 158–171
3 Goldman, A.I. (1989) Interpretation psychologized. *Mind Lang.* 4, 104–119
4 Astington, J.W. *et al.*, eds (1988) *Developing Theories of Mind*, Cambridge University Press
5 Wellman, H. (1990) *The Child's Theory of Mind*, MIT Press
6 Charness, G. and Rabin, M. (2002) Understanding social preferences with simple tests. *Q. J. Econ.* 117, 817–869
7 Fehr, E. and Schmidt, K.M. (2006) The economics of fairness, reciprocity and altruism – experimental evidence and new theories. In *Handbook of the Economics of Giving, Altruism and Reciprocity* (Vol. 1) (Kolm, S.C. and Ythier, J.M.,eds), pp. 615–691, North–Holland
8 Nash, J. (1951) Non-cooperative games. *Ann. Math.* 54, 286–295
9 Misyak, J.B. and Chater, N. (2014) Virtual bargaining: a theory of social decision-making. *Philos. Trans. R. Soc. B* (in press)
10 Grice, P. (1975) Logic and conversation. In *Syntax and Semantics 3: Speech Acts* (Cole, P. and Morgan, J., eds), pp. 41–58, Academic Press
11 Levinson, S. (2000) *Presumptive Meanings: The Theory of Generalized Conversational Implicature*, MIT Press
12 Sperber, D. and Wilson, D. (1986) *Relevance: Communication and Cognition*, Blackwell
13 Lewis, D. (1969) *Convention*, Harvard University Press
14 Aumann, R. (1976) Agreeing to disagree. *Ann. Stat.* 4, 1236–1239
15 Brown, P. and Levinson, S.C. (1987) *Politeness: Some Universals in Language Usage*, Cambridge University Press
16 Marsh, K.L. *et al.* (2009) Social connection through joint action and interpersonal coordination. *Top. Cogn. Sci.* 1, 320–339
17 Pickering, M.J. and Garrod, S. (2004) Toward a mechanistic psychology of dialogue. *Behav. Brain Sci.* 27, 169–189
18 Sebanz, N. *et al.* (2003) Representing others' actions: just like one's own? *Cognition* 88, 11–21
19 Kamm, F.M. (2007) *Intricate Ethics: Rights, Responsibilities, and Permissible Harms*, Oxford University Press
20 Lee, B.P.H. (2001) Mutual knowledge, background knowledge and shared beliefs: their roles in establishing common ground. *J. Pragmatics* 33, 21–44
21 Butterfill, S.A. and Sebanz, N. (2011) Joint action: what is shared? *Rev. Philos. Psychol.* 2, 137–146
22 Galantucci, B. and Sebanz, N. (2009) Joint action: current perspectives. *Top. Cogn. Sci.* 1, 255–259
23 Bratman, M.E. (2014) *Shared Agency: A Planning Theory of Acting Together*, Oxford University Press
24 Bernheim, B.D. (1984) Rationalizable strategic behavior. *Econometrica* 52, 1007–1028
25 Pearce, D.G. (1984) Rationalizable strategic behavior and the problem of perfection. *Econometrica* 52, 1029–1050
26 Nagel, R. (1995) Unraveling in guessing games: an experimental study. *Am. Econ. Rev.* 85, 1313–1326
27 Stahl, D.O. and Wilson, P.W. (1994) Experimental evidence on players' models of other players. *J. Econ. Behav. Organ.* 25, 309–327
28 Stahl, D.O. and Wilson, P.W. (1995) On players' models of other players: theory and experimental evidence. *Game Econ. Behav.* 10, 218–254
29 McKelvey, R.D. and Palfrey, T.R. (1995) Quantal response equilibria for normal form games. *Game Econ. Behav.* 10, 6–38
30 Bacharach, M. (1999) Interactive team reasoning: a contribution to the theory of co-operation. *Res. Econ.* 53, 117–147
31 Sugden, R. (2003) The logic of team reasoning. *Philos. Explor.* 6, 165–181
32 Butler, D.J. (2012) A choice for 'me' or for 'us'? Using we-reasoning to predict cooperation and coordination in games. *Theor. Decis.* 73, 53–76
33 Gallotti, M. and Frith, C.D. (2013) Social cognition in the we-mode. *Trends Cogn. Sci.* 17, 160–165
34 Schelling, T.C. (1960) *The Strategy of Conflict*, Harvard University Press
35 Sugden, R. (1995) A theory of focal points. *Econ. J.* 105, 533–550
36 Crawford, V.P. *et al.* (2008) The power of focal points is limited: even minute payoff asymmetry may yield large coordination failures. *Am. Econ. Rev.* 98, 1443–1458
37 Frank, M.C. and Goodman, N.D. (2012) Predicting pragmatic reasoning in language games. *Science* 336, 998
38 Bangerter, A. (2004) Using pointing and describing to achieve joint focus of attention in dialogue. *Psychol. Sci.* 15, 415–419
39 Behne, T. *et al.* (2012) Twelve-month-olds' comprehension and production of pointing. *Br. J. Dev. Psychol.* 30, 359–375
40 Moll, H. and Tomasello, M. (2004) 12- and 18-month-old infants follow gaze to spaces behind barriers. *Dev. Sci.* 7, F1–F9
41 Sacks, H. *et al.* (1974) A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735
42 Schegloff, E.A. (2000) Overlapping talk and the organization of turn-taking for conversation. *Lang. Soc.* 29, 1–63
43 Kendon, A. (1967) Some functions of gaze-direction in social interaction. *Acta Psychol.* 26, 22–63
44 Mazur, A. (1985) A biosocial model of status in face-to-face primate groups. *Soc. Forces* 64, 377–402
45 Heylen, D. (2006) Head gestures, gaze and the principles of conversational structure. *Int. J. Humanoid Robot.* 3, 1–27
46 Wagner, P. *et al.* (2014) Gesture and speech in interaction: an overview. *Speech Commun.* 57, 209–232
47 Gold, N. and Sugden, R. (2007) Collective intentions and team agency. *J. Philos.* 104, 109–137
48 Bratman, M.E. (1993) Shared intention. *Ethics* 104, 97–113
49 Gilbert, M. (2009) Shared intention and personal intentions. *Philos. Stud.* 144, 167–187
50 Searle, J.R. (1990) Collective intentions and actions. In *Intentions in Communication* (Cohen, P.R. *et al.*, eds), pp. 401–416, MIT Press
51 Sebanz, N. *et al.* (2006) Joint action: bodies and minds moving together. *Trends Cogn. Sci.* 10, 70–76
52 Knoblich, G. *et al.* (2011) Psychological research on joint action: theory and data. In *The Psychology of Learning and Motivation* (Vol. 54) (Ross, B., ed.), pp. 59–101, Academic Press
53 Hatfield, E. *et al.* (1993) Emotional contagion. *Curr. Dir. Psychol. Sci.* 2, 96–99
54 Epstein, R.A. (1999) Confusion about custom: disentangling informal customs from standard contractual provisions. *Univ. Chicago Law Rev.* 66, 821–835
55 Farrell, J. and Rabin, M. (1996) Cheap talk. *J. Econ. Perspect.* 10, 103–118