WISSENSCHAFTSZENTRUM BERLIN
FÜR SOZIALFORSCHUNG

SOCIAL SCIENCE RESEARCH
CENTER BERLIN

discussion papers

FS IV 98 - 11

Coordination and Learning with a Partial Language

Andreas Blume
University of Iowa

September 1998

ISSN Nr. 0722 - 6748

**Forschungsschwerpunkt
Marktprozeß und Unter-
nehmensentwicklung**

**Research Area
Market Processes and
Corporate Development**

Zitierweise/Citation:

Andreas Blume, **Coordination and Learning with a Partial Language,**
Discussion Paper FS IV 98 - 11, Wissenschaftszentrum Berlin, 1998.

Wissenschaftszentrum Berlin für Sozialforschung gGmbH,
Reichpietschufer 50, 10785 Berlin, Tel. (030) 2 54 91 - 0

ABSTRACT

Coordination and Learning with a Partial Language

by Andreas Blume*

This paper explores how efficiency promotes the use of structure in language. It starts from the premise that one of language's central characteristics is to provide a means for saying novel things about novel circumstances, its creativity. It is reasonable to expect that in a rich and changing environment, language will be incomplete. This encourages reliance on structure. It is shown how creative language use emerges form common knowledge structures, even if those structures are consistent with an *a priori* absence of a common language.

ZUSAMMENFASSUNG

Koordination und Lernen mit einer Partialsprache

In diesem Beitrag wird die Anwendung von Strukturen in einer Sprache aus Effizienzsicht begründet. Der Artikel geht davon aus, daß eines der wichtigsten Merkmale der Sprache in ihrer Kreativität zu sehen ist, d. h. als Mittel, um Neues über neue Sachverhalte auszusagen. Es ist deshalb zu erwarten, daß in einer vielfältigen und sich verändernden Umwelt die Sprache unvollständig bleiben wird. Dies fördert die Anwendung von Strukturen. Es wird gezeigt, wie die kreative Sprachanwendung aus allgemeinen Wissensstrukturen entsteht, auch dann, wenn diese Strukturen *a priori* noch keine gemeinsame Sprache bilden.

---

* This paper was begun while I enjoyed the hospitality of the MEDS department at Northwestern University. The first version was completed at the CentER for Economic Research at Tilburg University. I am grateful to David Austen-Smith, Eddie Dekel, Oguz Durumeric, Ehud Kalai, Bart Lipman, Alejandro Manelli and Roger Myerson for helpful discussions and to Vince Crawford, Hans Haller, Scott Page and Joel Sobel for comments. I have benefitted from comments by seminar participants at the CentER for Economic Research (Tilburg University), CREED (University of Amsterdam), University of Bielefeld, Humboldt University, Iowa State University, University of Saarbrücken, WZB (Social Science Research Center, Berlin), the SAET conference in Antalya, the Midwest Mathecon Conference at Indiana University, and the Summer Meeting at the University of València.

Drug firms need to get their inventions on to the market quickly. That is easier when researchers and factory designers talk to each other. ... As competition grows even more fierce, more companies may try to make sure that their boffins and factory designers talk to each other early on. For the money it saves, it might even be worth paying for an interpreter. [The Economist, November 9th 1996]

1 Introduction

This paper explores how efficiency promotes the use of structure in language. It starts from the premise that one of language's central characteristics is to provide a means for saying novel things about novel circumstances, its creativity (e.g. Aitchison [1993]). In a rich and changing environment, language will necessarily be incomplete. This encourages reliance on structure.

We will examine the role of structure in the use of a partial language and, more importantly, in learning a common language (i.e. in acquiring a complete lexicon) from a partial language. Structure facilitates some coordination tasks and may result in *fast learning.*

While we use the term "language" in a wider sense than "natural language," natural language does have structure that both enhances its usefulness and learnability. Prominent examples are modularity (as appears in the distinction of verbs and nouns, of word stem and suffix, etc.) and the use of (spatial, temporal, causal, ...) order. Meaningful communication depends on contextual information that has to be learned in any new situation. The sentence "Let's meet at 10:00 a.m. at the Rhetoric Building," and many sentence like it, become meaningful only once a particular building is specified (labeled) as the Rhetoric building. The process by which the sentence acquires meaning relies both on the modular *structure* of natural language and a shared understanding of a *labeling rule.* Thus, new situations are characterized by languages that leave at least some objects unlabelled, and learning a common language requires that these objects become labelled. In this paper, we will emphasize this kind of novelty, and consider a world in which labels for individual objects are entirely missing. The aim is to develop a coherent framework for the evolution of language when agents lack a common language, but have common knowledge of a structure and of a labeling rule.

Formally, a language will be a collection of rankings (or labelings) of a finite set of objects. If this collection is a singleton, then we say that there is a *common language.* In this case, every object is labeled unambiguously. If there are multiple rankings, there will be alternative labelings for at least some of the objects. To highlight the structural aspects of language, we concentrate on the case where a language does not make any distinctions among individual objects. Even in this case of *absence of a common language* there will often be a partial language that facilitates coordination on sets of objects and/or permits a common language to be learned quickly.

Two simple examples illustrate the role of a partial language for coordination and learning. First, consider the problem of two agents trying to coordinate on one of two projects and at the same time on the assignment of one of two tasks within each project.[1] Suppose that payoffs are positive if and only if both agents pick the same project and different tasks. If there is symmetry among projects and among tasks, so that agents have no obvious reason to choose one project (or task) over another, then one can argue that the probability of coordination is one fourth. Now suppose that in addition each agent has the option to divide her efforts among two project-task combinations (she could for example pick one project and perform both tasks). Let payoffs be positive if and only if for each project-task choice made by one agent, the other agent makes a complementary choice (so, if one agent performs the first task for the first project and the second task for the second project, then the other agent must perform the second task for the first project and the first task for the second project). Note that formally this game is equivalent to one in which agents choose one- or two-element subsets from a four-element set with the objective of choosing identical sets. Despite the symmetry of projects and of tasks, and even if the agents perceive the categories "projects" and "tasks" symmetrically, there is not complete symmetry among project-task pairs. There are only two pairs that include both projects and both tasks, while there are four pairs that combine both choices in one category with one in the other. If agents make use of this asymmetry, they can raise the coordination probability to one half.

Second, consider a problem of tacit collusion. Let there be a stream of public projects (one per period) that a small number, n, of firms bid on. The firms may want to establish some form of bidding rotation.[2]

Without obvious criteria by which to rank firms or with too many conflicting criteria, firms will be symmetric, and establishing the order of rotation is difficult. Symmetry (i.e. indistinguishability) among firms becomes

[1] There is an obvious generalization to any number of objects, tasks and agents.

[2] A classical example of the use of an elaborate bidding rotation is provided by the "electric equipment conspiracies" of the 1950's (Fuller [1962]). The firms involved made use of the phases of the moon to coordinate their bids. In this case collusion was achieved via explicit agreement.

a source of strategic uncertainty and the firms have to rely on the history of their strategic interaction to break this symmetry.

Suppose that firms learn optimally in the sense that their strategies are efficient subject to the constraint that after every history strategies respect the remaining symmetries. Then collusion will be achieved as soon as all symmetries have been removed. The time it takes to coordinate will then depend on the initial degree of symmetry and on the information that becomes available in each period's strategic interaction.

To appreciate the role of information, note that if all bids are revealed, a single observation may suffice to distinguish all firms and thereby to establish collusion. If, in contrast, firms are completely symmetric and only the winning bid can be observed after each round, then a minimum of $n-1$ observations is needed to achieve collusion.

To appreciate the role of structure, continue with the assumption that only the winning bid is observed and consider the following three cases: (1) firms can naturally be split into two equal sized groups, with no obvious ranking of the groups or within the groups, (2) firms can be split into two groups and the groups can be ranked, but firms within a group cannot, and (3) there has been a history of rotation that was somehow interrupted (e.g. through entry and subsequent exit of rival firms) leaving no indication whose turn it is next.

In none of the cases can collusion be guaranteed *a priori.* In the first case $n-2$ observations are needed to achieve collusion. The number of observations needed is the same in the second case, except that here first-period competition can be reduced by having only firms in the first group bid. In the final case, while the firms are *a priori* indistinguishable, it takes only one observation to reinitiate the rotation.

In both examples the use of the available structure arises out of efficient equilibrium play subject to symmetry constraints. Following Crawford and Haller [1990] (CH in the sequel), strategies that respect at each point in time whatever symmetries remain in a game will be called *attainable.* The efficient strategies in this class will be referred to as *efficient attainable strategies,* and if the game is a coordination game, as *optimal attainable strategies.* In a repeated coordination game environment, the use of optimal attainable strategies is a form of *optimal learning* (to distinguish a priori identical objects).

Two interpretations for the focus on optimal attainable strategies are available, an evolutionary or learning interpretation and a mechanism design interpretation.[3] Under the former there is evolutionary pressure on behaviors that depend on factors that are constant across novel circumstances. The mechanism design interpretation on the other hand asks directly for learning rules that cope efficiently with novelty.

We follow CH's approach of modeling alternative descriptions of a set of objects as permutations of some ranking of the set. Players' ignorance about other players' descriptions is then expressed by their beliefs over the set of alternative permutations. This approach permits us to model a partial language in a standard Bayesian game setting.

CH considered the case where the set of descriptions consisted of all permutations.[4] We depart from CH by permitting sets of descriptions to be nontrivial subsets of the set of all permutations.

For the set of descriptions to be common knowledge among players it must be the case that players are unable to arrive at a simpler set via introspection. There must not for example be a single description in the set of descriptions that is distinct from all other members of the set. One requirement for this to hold is uniformity of beliefs. Another is irreducibility of the set of descriptions. We will show that irreducibility amounts to the set of descriptions forming a subgroup (in the mathematical sense) of the group of all permutations. This is quite natural given that irreducibility expresses symmetry among the permutations in the set; no subset can be singled out.

The next section formalizes the notion of attainable strategies, compares our approach to CH's, and gives examples for creative language use, fast learning and the role of subsets of the set of all permutations. In Section 3 we show that an irreducible set corresponds to a permutation group and that absence of a common description is achieved by a transitive subgroup. In Section 4 we show that even with absence of a common description a partial language may aid certain coordination tasks. In Section 5 we show that

[3] Rubinstein [1996] calls the designer of a such a mechanism a "linguistic engineer."

[4] This is the assumption used in the body of CH's paper. In the appendix they consider a more general formulation that is compatible with the approach in the present paper. There the extent of a common description of a set of objects is expressed via a partition of that set, with absence of a common description captured by the trivial partition. The partition is refined as a function of the history of play, where a very general form of history dependence is possible. CH do not closely investigate the relation between histories and partitions. For example, for two objects not to be commonly distinguished it must be the case that the partitions they induce, if we exchange them in otherwise identical histories, must be identical up to symmetry. The approach of the present paper ensures that this is the case.

learning preserves irreducibility, that absence of a common description is compatible with fast learning and that, in a sense, fast learning is ubiquitous. Section 6 explores the variety of fast learning phenomena when there is absence of a common description. Section 7 discusses the related literature. Section 8 concludes.

2 Strategies, Games and Examples

This section recalls CH's definition of an *attainable strategy*, compares our approach to CH's, and applies the concept of attainable strategy to settings with creative language use and fast learning.

Let - denote a finite set of objects for which players in a game lack a common language. As part of the description of a game, - might represent actions of players, some part of their private information, or even the players themselves. Strategies, as usual, are functions that map the players' information into their actions. The fact that players lack a common language description of - , formally appears as a restriction on their beliefs.

This approach permits us to model the absence of a common description in a standard Bayesian game setting. Players' descriptions are drawn from a common-knowledge distribution ϕ on the set of possible descriptions of - , where we take this set as the set of permutations of the elements of - . The description drawn by a player becomes that players' private information.

A player's strategy may well depend in great detail on the set - . However from the perspective of the other players it can only depend on those aspects of - that have a common description. If therefore we adopt the convention that a player's strategy expresses the beliefs other players hold about her (see for example Aumann and Brandenburger [1995]), then lack of common describability can be expressed in terms of restrictions on players' strategies. Therefore we call a strategy *attainable* for a player if it satisfies the condition that any two pure strategies that differ only in the treatment of elements of - that are not commonly distinguished are equally likely.

When considering coordination games, we are interested in those attainable strategies that maximize the players' ex ante payoff. These we call *optimal attainable strategies;* in more general settings, one would look for efficient attainable strategies. Note that, like CH, we deliberately ignore higher order coordination problems that may arise if there are multiple optimal attainable strategies. In the spirit of the introduction we think of those multiplicities as being eliminated either by evolutionary pressures or by Rubinstein's [1996] "linguistic engineer."

Next we discuss two examples to illustrate the role of optimal attainable strategies. The first example recalls CH's analysis and suggests how it might be altered if players have a partial language. The second example illustrates creative language use in a simple communication game and the multiple roles of the set - in different environments.

2.1 Repeated Coordination Games

Here we recapitulate a central insight from the work of Crawford and Haller and suggest how additional structure may lead to fast learning. Consider the following game played on a finite set of objects - : Each of two players chooses one element of - , one object, simultaneously and independently. If both make identical choices, then their payoffs equal 1, otherwise their payoffs are equal to 0. Let this game be repeated infinitely often, and let repeated game payoffs be equal to the discounted sum of stage game payoffs, with a discount factor $0 < \delta < 1$. Let the players' uncertainty be represented by a uniform distribution over all possible permutations of some ranking of - . Also, let the players' positions be not distinguished, which forces them to use identical strategies.

CH show that for #- $= 2$, there is an essentially unique optimal attainable strategy, according to which players randomize uniformly over the two objects until they achieve a match, and then stick to the matching object. With #- ≥ 6 there is an optimal attainable strategy that requires uniform randomization over all objects in the first period, and thereafter continuation with the same object as soon as there is a match, and uniform randomization over the two objects chosen in the first period as long as there was no match.

Note that the expected coordination time is $\hat{t} \geq 2$, that with positive probability coordination will take more than two periods, and that even if coordination is achieved, the players are in general far from having developed a common language. In other contexts, as in our tacit collusion example, learning a common language fully may be essential.

For a simple geometric example of fast learning, replace - by the set of points on a sphere (with no further distinctions among those points). Now the probability of coordination in the first round equals zero. However, the strategy of picking the midpoint of the shortest distance between the two first-round choices is attainable, and with probability one guarantees coordination in the second round. In this case fast learning results because acknowledging that locations are arranged on a sphere implicitly limits the set of permutations to motions, i.e. all those permutations of points on the sphere that leave distances invariant.

2.2 A Rudimentary Grammar

Consider the following game played repeatedly between two players, a sender and a receiver. At the beginning of the game the sender learns his private information and sends a message to the receiver. Upon receiving the message the receiver takes an action. Payoffs depend only on the sender's private information, his type, and the receiver's action.[5] The payoff to both players is one if the receiver's action matches the sender's type and zero otherwise.[6] There is exactly one matching action for each type of the sender. Messages do not affect payoffs directly. Assume that after each round the players commonly observe the type drawn for that round, the message sent in that round, and the action taken. The sender's private information is determined anew in each round according to a uniform distribution.

Additional structure is provided by types and messages being strings. To simplify the discussion let types be triples formed by rearrangements of the letters A, B and C, e.g. (B, C, A), and let messages be triples as well, formed by rearrangements of $*, \#$ and &. To rule out *a priori* focal points we will assume that each players has a private description of these symbols.[7]

The first time the game is played, the sender and the receiver lack a common-knowledge description of the types space and of the message space.[8] According to our definition of attainable strategy, each type randomizes uniformly over all messages and the probability of a matching action being taken is $1/6$. In the following round however each player has observed two triples that can be used to construct a bijection between type symbols and message symbols. Using this bijection, every type can be uniquely identified with a message. This means that there is an attainable strategy guaranteeing successful communication in the second round and thereafter.

[5]Wärneryd [1993], Blume, Kim and Sobel [1993] and Blume [1996] consider the evolution of meaning of *a priori* meaningless messages in sender-receiver games.

[6]If one is willing to consider efficient attainable strategies, the example generalizes to arbitrary incentive structures.

[7]The focal point notion was first proposed by Schelling [1960]. See Blume, DeJong, Kim and Sprinkle [1998] for an example of the inducement of private descriptions in an experimental setting.

[8]It would suffice for the example if we let only the type space have a private description.

Contrary to the first example, coordination is guaranteed from the second period on and a common language is learned. Moreover players use their language creatively in that in the second period they are likely to indicate a novel type (not observed before) via a novel message (not sent before). Batali [1996] has referred to similar structures as *grammars,* emphasized the role of such grammars for the expression of novel meanings and inquired into the evolution of such grammars.

3 Partial Languages

In this section, we will formally define a language, discuss irreducibility of a language, and say what we mean by absence of a common language. A language will be a set of rankings of a finite set of objects. Irreducibility will be seen to imply that the language induces a group action on the set of objects, and absence of a common language will be captured by this group action being transitive.

Given a finite set - of objects, $\#(-) = n$, we define a *language* to be a set of rankings, R, of the elements of - . Thus, a language is the set of labelings of - that agents consider possible. We will assume that the language is common knowledge among the players.[9] Common knowledge of the language requires that there is no structure in R that identifies an alternative set of rankings that is in some way more effective (e.g. makes more distinctions among objects, or permits a common language to be learned faster) than R. Otherwise players could, at least in principle, rely on introspection to adopt the more effective language. A language R must be *internally irreducible,* i.e. there must be symmetry among the elements of R that prevents a nontrivial subset of R from being singled out. The language must also be *externally irreducible,* such that any ranking outside of R belongs to a language $\tilde{R}$ that is essentially equivalent to R.

We will formalize internal irreducibility; it turns out that internal ir-

[9]This is a simplifying assumption. Alternatively, one could examine situations in which different agents entertain different languages, e.g. one agent may be capable of making finer distinctions among objects than another agent. While this is beyond the scope of the present paper, it should be noted that Bacharach's [1993] theory of "variable universe games" seems to be an appropriate framework for modelling multiple possible languages, or variable frames. In particular, Bacharach addresses the problem of players' beliefs about each other's languages. One possibility is for example, that the beliefs of a player with a relatively coarse language do not put any weight on another player having a finer language. Bacharach and Bernasconi [1997] test variable frame theory experimentally.

reducibility of a language implies a natural form of external irreducibility. Note that if we pick any ranking r in R as the standard in which we enumerate - , then the others can be expressed as permutations of r. For R to be symmetric, the resulting set of permutations must be independent of the choice of standard $r \in R$. For any given r, let H denote the set of permutations that generate R, i.e. $R = Hr$. Let r' be another element of R and let $g \in H$ be such that

$$g(r(\omega)) = r'(\omega) \ \forall \omega \in - .$$

Then

$$R = Hr = Hg^{-1}r'.$$

Therefore our desired internal irreducibility condition is

$$Hg^{-1} = H \ \forall g \in H.$$

One easily checks that this condition is equivalent to H being a subgroup of the symmetric group S of all permutations of the set - . The group H *acts* naturally on the set of objects - .[10]

Analogously, for any two rankings $\tilde{r}$ and $\hat{r}$ that are not in R and with $\tilde{g}(r) = \tilde{r}$ and $\hat{g}(r) = \hat{r}$, we have

$$R = Hr = H\tilde{g}^{-1}\tilde{r} = H\hat{g}^{-1}\hat{r}.$$

Therefore $\tilde{r}$ and $\hat{r}$ are indistinguishable in terms of R if

$$H\tilde{g}^{-1} = H\hat{g}^{-1}.$$

If H forms a group, this is equivalent to

$$\tilde{g} \in \hat{g}H.$$

This is a (left-) coset of H and it is well-known that the cosets of H partition S . Therefore, if H is a subgroup, we can partition the set of all possible rankings into subsets that are permutations of R, and as a consequence internal irreducibility implies external irreducibility.

To summarize, if R induces a subgroup H of S , then the resulting symmetries prevent agents from coordinating on any set of rankings that is simpler than R. The elements of R are all symmetric to each other, and for

[10] For convenience, a few simple facts about groups are collected in the appendix. Rotman [1996] gives an elementary introduction to groups, and Dixon and Mortimer [1996] offer an up-to-date account of permutation groups.

any ranking $\tilde{r}$ outside of R, there exists a set of rankings that is symmetric to R and contains $\tilde{r}$. From now on we will simply identify a language with the corresponding group. A partial language is then a nontrivial proper subgroup of S , and a common language the trivial subgroup consisting of the group identity alone.

Having resolved the issue of irreducibility of a language, we now turn to the question of how a language distinguishes among objects. The answer is that two objects are distinguished by a language if and only if the set of labels assigned to the objects differ. This can be formalized in terms of the action of the group H (that represents the language) on the set of objects, - .

Given any group G acting on - , we can associate with any $\omega \in$ - the image of ω under the group action. Define the *orbit* of ω under G acting on - as

$$\mathcal{O}(\omega) := \{g(\omega) | g \in G\} \subset - .$$

Then points ω and ω' can be distinguished if and only if they belong to different orbits. In that regard, it is useful to know that the set of orbits forms a partition of - (e.g. Rotman [1996], p.122).

For any partial language, there will be at least one nontrivial subset of the set of objects among which the agents cannot make common distinctions. For any such subset the agents lack a common description, which limits their ability to coordinate and forces them to learn to make such distinctions. Since it is in the nature of a partial language that there is *some* lack of a common description, we will focus on this aspect and for the most part assume that agents can make *no* common distinctions among individual objects.

We focus on the case of absence of a common description to emphasize the role of structure in a partial language, as a benchmark and to emphasize the contrast with CH. Formally, we have absence of a common description when the language induces a single orbit. Then $\mathcal{O}(\omega) =$ - $\forall \omega \in$ - . In this instance none of the elements of - can be distinguished from any other element of - . A group H, acting on - , with only a single orbit is called *transitive* (e.g. Dixon and Mortimer [1996], p.8). Thus transitivity of the group action formalizes our intuitive notion of individual objects lacking common distinctions. Note that even if a partial language makes some distinctions among individual objects, i.e. has multiple orbits, the group representing it acts transitively on each of its orbits; therefore transitivity will be an issue in any partial language.[11]

As an example, consider the following language, given by a set of six possible rankings, corresponding to the columns below, of nine objects, corresponding to the rows.

1	4	7	1	4	7
2	5	8	3	6	9
3	6	9	2	5	8
4	7	1	4	7	1
5	8	2	6	9	3
6	9	3	5	8	2
7	1	4	7	1	4
8	2	5	9	3	6
9	3	6	8	2	5

This language induces two orbits and thus distinguishes the set of objects with the potential labels 1,4 and 7 from the set that is associated with the remaining six labels. Thus, in terms of the entire set, we do not have absence of a common description. However, the inability to distinguish the elements that belong to the same orbit is due to absence of a common description for those elements. For future reference, note that labeling one of the objects in the first set implicitly singles out a pair of rankings, namely the rankings consistent with that labeling, whereas labeling one of the objects in the second set identifies exactly one ranking.

4 Coordination with a Partial Language

A partial language that makes no distinctions among individual objects may still aid certain coordination tasks.[12] In the introduction we provided an example that involved multiple, cross categorizations of objects in a set. In that environment it was easier to coordinate on a nontrivial subset of the set of objects than on an individual object, even though there are many more subsets than objects. In this section we generalize this idea.

The usefulness of a partial language G on - in a game depends on the role played by - in that game. For example, we can think of (1) a one-shot

[11] If we are interested in the role of structure in language in promoting coordination and learning, then by restricting attention to the case of absence of a common description, we are tying our hands. We are ruling out the obvious benefits from having a language that already makes some (even if only coarse) distinctions at the level of individual objects.

[12] I am grateful to a referee for alerting me to this fact.

coordination game where the actions are single elements of Ω, (2) a repeated coordination game with the same action set, or (3) a one-shot coordination game where the actions are m-element subsets of Ω. A partial language that is useless in (1) may still be useful in (2) or (3). Multiple categorizations, as in our introductory example, are one source of partial languages with this property. An ordering of the elements of Ω is another, provided there is no maximal or minimal element, which in the finite case amounts to the ordering being cyclic.

For example, let G be the cyclic group generated by an n-cycle $g \in S_{\Omega}$. Then G acts transitively on Ω and thus individual objects are indistinguishable. However, G acts also on the set $\Omega^{\{2\}}$ of 2-element subsets of Ω. The action of G on $\Omega^{\{2\}}$ is by no means transitive. For example the orbit of an adjacent pair consists entirely of adjacent pairs. Note that there are just as many adjacent pairs as there are elements of Ω. Therefore, with this partial language G coordination on 2-sets is no more difficult than coordination on individual objects, even though $\#(\Omega^{\{2\}}) = \begin{pmatrix} n \\ 2 \end{pmatrix} > n = \#(\Omega)$ for $n > 3$. Indeed, in the case where n is even, coordination on 2-sets is easier than coordination on singletons since there are only $\frac{n}{2}$ antipodal pairs. This observation can be generalized by asking when it is the case that coordination on a k-set with $k < n$ is easier than coordination on a singleton.

Let $\Omega^{\{k\}}$ denote the set of all k-element subsets of Ω, and let $x \in \Omega^{\{k\}}$. The difficulty of coordinating on $\Omega^{\{k\}}$ is then measured by the size of the minimal orbit of G acting on $\Omega^{\{k\}}$. We want to know which k minimizes the difficulty of coordination. To answer this question, it helps to note a few simple properties of orbits and stabilizers: For any $x \in \Omega^{\{k\}}$, we have

$$\#(\mathcal{O}(x)) = \frac{|G|}{|G_x|},$$

which is known as the *orbit-stabilizer property* and is a simple consequence of Lagrange's theorem. It follows that $|G_x|$ and $\mathcal{O}(x)$ are divisors of $|G|$. Moreover, if G is generated by an n-cycle and $k < n$, we have $|G| = n$ and $|G_x| < n$. Therefore the smallest orbit cannot be smaller than the smallest prime divisor of n, denoted by $p(n)$. To see that in fact the size of the smallest orbit is equal to $p(n)$, consider the set

$$\{\omega, g^{p(n)}(\omega), g^{2p(n)}(\omega) \ldots\}.$$

The orbit of this set has exactly $p(n)$ elements. Thus we have the following observation.

Proposition 1 *If the language G on - , with $\#(-) = n$, is generated by an n-cycle and $p(n)$ is the smallest prime divisor of n, then among all k-sets with $k < n$ it is easiest to coordinate on one that has $k = \frac{n}{p(n)}$ elements.*

Thus far we have shown that sometimes there are partial languages for which players are better off choosing k-sets with $k > 1$ than choosing singletons. Partial languages do not always help. If G is for example the alternating group of degree four, then coordination on any k-set is just as hard as without any language at all. There is however a weak general result that guarantees a role for a partial language in coordination on k-sets if that language helps with coordination on sets of lesser size, as long as $k < n/2$. Say that G "facilitates coordination on m-sets" if G has multiple orbits on $-^{\{m\}}$.

Proposition 2 *If G is a language on - , then for $0 \leq m \leq k$, and $m + k \leq \#(-)$, if G facilitates coordination on m-sets, it does so on k-sets.*

Proof. This result is an immediate consequence of Theorem 9.4A in Dixon and Mortimer [1996], who also provide references for the history of this result. □

5 Learning to Coordinate with a Partial Language

In this section, we investigate how agents use observations of sets of objects to learn a common language from a partial language. One application is learning to coordinate in repeated two-player coordination games when players share a partial language on the set of actions - . This generalizes the work of Crawford and Haller, who studied the case of learning without any *a priori* language.

In a dynamic setting players can use their observations of objects to label these objects and thus refine their language. At a minimum, the observed objects themselves become distinct from the others. In general however, with a partial language H already in place, the observation of a set of objects induces further distinctions. This is the case in the introductory tacit-collusion example: With a history of rotation, and without knowing where to start, a single observation of a winning bid suffices to label all firms, not just the winner.

In addition to common knowledge of a structure on the set of objects, we will require common knowledge of a *labeling rule.* In the case of singleton observations, a plausible labeling rule would be to assign to each newly observed object the lowest remaining number that is available for that object given the partial language at that point in time. This just uses the ordering of the natural numbers. More generally, we will assume that there are commonly known orderings of the sets of subsets of $\{1, \ldots, n\}$ of a given size, e.g. an ordering of all the pairs, an ordering of all the triples, etc.. A possible labeling rule based on such a collection of orderings would assign to any observed set Δ the lowest numbered set of labels that remain available for Δ, according to the partial language at that point in time.[13]

As an example, let there be n objects $n-3$ of which have already been labeled at time t, and the partial language that prevails at time t permits all possible assignments of the labels 3, 7, and 9 to the remaining three objects. If a pair of unlabeled objects is observed, the labeling rule determines a set of labels out of $\{3,7\}$, $\{3,9\}$, and $\{7,9\}$, e.g. $\{3,7\}$. As a result the one unobserved object receives the label 9, and there is a pair of objects, for which it is not yet known which object receives the label 3 and which the label 7.

Besides labeling objects, observations distinguish rankings. In the example there are two rankings that permit the set of labels $\{3,7\}$ for the observed pair of objects. If in addition the partial language at time t expresses a circular order on the three remaining unobserved objects (while making no distinctions among these objects), then there is only one ranking that permits the labels 3 and 7 for the observed pair.

In general, if a set Δ is observed, the labeling rule determines not only which labels to assign to Δ but also singles out the rankings that label Δ in this way. If we make any of these rankings the standard, this amounts to selecting those permutations that fix Δ. Therefore the evolution of a language due to observations can be expressed (a) in terms of the labeling of observed sets of objects, and (b) the selection of permutations that in an appropriate sense fix the observed sets of objects.

The elements of H which are thus identified by an observation $\Delta \subset$ -

[13] Common knowledge of a labeling rule is a further simplifying assumption, in addition to common knowledge of structure on the set of objects. These assumptions help us to focus on one type of constraint on the learning process, namely novelty of individual objects. The assumption of common knowledge of a labeling rule is more plausible in some contexts than in others. It is implicit in CH's focus on optimal attainable strategies, when they assume that following coordination, players will stick to the coordinating action, rather than switch to another uniquely distinguished action. There are other instances where a particular rule may be focal. For example, the rule proposed for singleton observations does appear prominent.

are the *setwise stabilizer* $H_{\{\Delta\}}$ of Δ. It is well known that if H is a group, then for any Δ, the corresponding setwise stabilizer is a subgroup of H. Therefore we have the following observation.

Proposition 3 *Given a language* $H \leq S$ *, an observation* $\Delta \subset$ - *preserves irreducibility by inducing a language* $H_{\{\Delta\}}$.

Thus, learning a language amounts to using observations to construct ever smaller subgroups from a given group, or partial language. If this sequence converges to the group identity, we say that a common language has been learned.

At this point one might ask whether agents might prefer to update the support of their beliefs based on - $\setminus \Delta$ rather than Δ. For that purpose note that when a group G acts on - , then each function $g(\cdot)$ is a bijection since its inverse is simply $g^{-1}(\cdot)$, where g^{-1} is the inverse of g in the group G. It is easily seen that if each $g(\cdot)$ is a bijection, then

$$H_{\{\Delta\}} = H_{\{\Delta^C\}} \ \ \forall \Delta \subset - .$$

Thus, observing a set is equivalent to observing its complement.[14]

Next, we reexamine the question of how hard it is to solve two-player coordination problems when there is absence of a common language. CH focussed on the case where the players' initial beliefs take the form of a uniform distribution over all possible permutations.[15] Then each observation identifies merely the observation itself, and does not lead to any further differentiation of the set - . Moreover, once coordination is achieved, at most three sets of actions are identified, the coordinating action, the unsuccessful action and the unused actions.Thus the optimal coordinating process is limited in scope. No common distinction among unused messages arises; players do not use their language creatively.

Our examples of a rudimentary grammar and coordination on spheres show that absence of a common description is consistent with rapid coordination and learning of a common language. Our next result examines the

[14]This also explains why Crawford and Haller note that in coordination games with three actions, and without any prior language, if players do not coordinate in the first round, then it is optimal for them to continue with the one unused action. The observation of the set of two actions which did not lead to coordination is equivalent to an observation of the complement of that set (I owe this observation to Scott Page). Blume and Gneezy [1998] find some experimental evidence for this form of learning in three-location games.

[15]In an appendix they consider a general formulation using partitions but do not investigate optimal learning in this case.

intuition underlying these examples formally, in a finite setting. Since the result deals only with n-cycles on - , it is worth noting that cycles are the natural representation of order in the finite setting when there is no maximal (minimal) element.[16] Furthermore, it is straightforward to use cycles as building blocks for a whole panoply of partial languages (be they transitive or not).

When positions in a two-player simultaneous-move game are not distinguished, then observations of the two players' simultaneous actions are not distinguished and therefore we must consider the corresponding setwise stabilizers. We will refer to observations of k-element sets as "k-observations." The next result on the effect of two-observations is applicable to two-player games in which players' positions are not distinguished; the generalization to more than two players is straightforward.

Proposition 4 *Let $G =< g >$ where g is an n-cycle and consider the natural action of G on - . Then*

1. *G expresses absence of a common description;*

2. *if n is odd, then every two-observation induces a common description of all elements of - ; and,*

3. *if n is even, then a proportion $\frac{n}{n+1}$ of all possible two-observations induce a common description of all elements of - .*

Proof. For (1) to hold, G must have a single orbit. The definition of an n-cycle implies immediately that G does indeed have a single orbit.

For (2) let n be odd and consider, without loss of generality, the two-observation $\{\omega_1, \omega_2\}$. Since g is an n-cycle, e is the unique element $h \in G$ such that $h(\omega_1) = \omega_1$. To derive a contradiction, suppose that $G_{\{\{\omega_1,\omega_2\}\}}$ is not a singleton. Then there must be numbers k, l, $0 \leq k, l \leq n-1$ such that

$$g^k(\omega_1) = g^l(\omega_2),$$

and

$$g^k(\omega_2) = g^l(\omega_1),$$

which can be restated as $g^{k-l}(\omega_1) = \omega_2$ and $g^{k-l}(\omega_2) = \omega_1$. Therefore $g^{2(k-l)}(\omega_2) = \omega_2$. Let $k \geq l$ without loss of generality. Then, since g is an n-cycle by assumption, n must be a divisor of $2(k-l)$ and since $2(k-l) < 2n$,

[16] See Rubinstein [1996] for the role of order in natural language.

it must be the case that $2(k-l)=n$. This implies that n is even, thus generating a contradiction. Therefore $G_{\{\{\omega_1,\omega_2\}\}}$ must be a singleton.

The argument that we just gave to show that $G_{\{\{\omega_1,\omega_2\}\}}$ must be a singleton whenever n is odd, clearly does not work for the case where n is even, and indeed the claim is not true when n is even. However, it remains true that any two-observation $\{\omega_1,\omega_2\}$ whose stabilizer is not a singleton satisfies

$$g^{\frac{n}{2}}(\omega_1)=\omega_2.$$

There are $\frac{n}{2}$ such two-observations. The total number of two-observations is $n+\frac{n(n-1)}{2}$, consisting of n pairs with identical observations, and $\frac{n(n-1)}{2}$ pairs with distinct observations (where we divide by 2 because order does not matter). Therefore the proportion of two-observations which lead to a common description is

$$\frac{n+\frac{n(n-1)}{2}-\frac{n}{2}}{n+\frac{n(n-1)}{2}}=\frac{n}{n+1}.$$

□

Consider for example four locations that are positioned symmetrically on a wheel that has a direction. If two players choose locations independently, there are three types of choice pairs that can arise: four pairs with identical choices, four pairs with adjacent choices, and two pairs with antipodal choices. Given the direction on the wheel, there exists a labeling rule (e.g. assign the label "1" to the left element of each pair, where possible, and label the remaining locations consecutively) that identifies a common language for all two-observations that are not antipodal. The proportion of two-observations that are not antipodal is 4/5.

When positions are distinguished in a two-player simultaneous-move game, the two players' simultaneous actions become separate singleton observations. If the game in question is a simple coordination game and we are interested in how players achieve coordination through repeated play, then learning with distinguished positions is essentially trivial; since positions are distinguished, we can simply assign the (commonly known) label "player one" to one of the players. There is then an attainable strategy profile in which player one repeats his first-round action in subsequent rounds and player two uses a best reply to player one's action in all rounds following the first round. This is independent of which subgroup of the symmetric group S expresses the players' uncertainty at the beginning of the repeated game.

Learning with distinguished positions becomes less trivial if we alter the game. For example, consider a game in which, as before, players receive a positive payoff if and only if they meet in some location but once a locations has served as a meeting place it cannot do so again until at least k rounds have passed, where $1 < k \leq n-1$. Also, simplify by letting only player one's action be commonly observable after each round. If the initial uncertainty is described by S , then locations become identified only by player one's choice of those locations. Successful coordination on some location in the first period, for example, does not guarantee coordination in subsequent periods because that location cannot be revisited for some time and because for the other locations any kind of common description is still lacking.

Especially for large n and k, coordination in the initial phase of the game becomes quite tedious. Consider as an extreme case $k = n - 1$. Then sustained coordination is possible but in order to achieve it, the players need to acquire a complete common description of the set of locations - . Even if the players cared about nothing else but achieving such a common description as early as possible, it would still take $n-1$ periods. Given, that players discount the future and given that coordination is a chance event at any location that is not yet commonly described, the expected time until a full common description is achieved exceeds $n - 1$. For example, if agents do not coordinate in the first period, then discounting will induce them to both visit player one's first period choice in the second period. Thus, no new location is identified in the second period.

In this example the full description of - is acquired very slowly, one observation at a time. This contrasts with the case where the initial uncertainty is described by $< (\omega_1, \omega_2, \ldots, \omega_n) >$ and where therefore coordination can be guaranteed in all rounds but the first round.

For the remainder of the paper we will concentrate on the case of singleton observations, as in the example with distinguished positions. For the next result we will also focus on essential observations, that is ignore observations of objects that are already labeled. We are interested in how many essential observations are needed to learn a common language.

Definition 1 *Given a set of observations $\Delta' \subset$ - , an observation $\omega \in$ - is essential if $G_{(\Delta')} \neq G_{(\Delta' \cup \{\omega)\}}$. A set of observations Δ is essential if it can be arranged in an order $\omega_1, \ldots, \omega_k$, such that ω_j is essential given $\{\omega_1, \ldots, \omega_{j-1}\}$ for all $j = 1, \ldots, k$.*

Definition 2 *Let G, a partial language on - , be given. Then* the maximum number of essential observations needed to learn a common language *is equal to k if*

1. *for any essential set* $\Delta \subset$ - , *with* $\#(\Delta) = k$, *the pointwise stabilizer* $G_{(\Delta)}$ *equals the group identity, and*

2. *there exists an essential set* $\Delta' \subset$ - , $\#(\Delta') = k - 1$, *such that* $G_{(\Delta')}$ *differs from the group identity.*

In the example at the end of section 3 there are three kinds of essential sets, singleton sets that identify a common language (one of the objects associated with the labels 2, 3, 5, 6, 8 and 9), singleton sets that identify a pair of possible common languages, and pairs of objects that identify a common language. Hence, in the example, the maximum number of essential observations needed to learn a common language equals two, even though a single observation of the right kind would suffice to learn a common language.

Obviously, for $\#($- $) = n$ the maximum number of essential observations needed to learn a common language never exceeds $n - 1$. Our previous examples show that it may be smaller. Indeed, the following result shows that the case where $n - 1$ essential observations are needed is in a certain sense atypical. For any partial language that corresponds to a proper subgroup G of S , denoted $G < S$, strictly less than $n - 1$ essential observations are needed to learn a common language.

Proposition 5 *For any partial language* $G < S$ *on* - , *with* $\#($- $) = n$, *the maximum number of essential observations needed to learn a common language is strictly less than* $n - 1$.

Proof. If G is not transitive on - , then we can partition - into the orbits of G on - (e.g. Rotman [1996], p.122). Let there be $r > 1$ such orbits and denote the $\rho's$ orbit by $\mathcal{O}_\rho$. Then $n = \sum_{\rho=1}^{r} \#(\mathcal{O}_\rho)$. Any essential set includes at most $\#(\mathcal{O}_\rho) - 1$ observations from the ρ's orbit. Therefore the cardinality of an essential set cannot exceed $\sum_{\rho=1}^{r}(\#(\mathcal{O}_\rho) - 1) = n - r$ observations. It follows that the maximum number of essential observations needed to learn a common language is no larger than $n - r$. The same argument works if for some $r > 1$ and for every set Δ of $n - r$ essential observations the pointwise stabilizer $G_{(\Delta)}$ fails to be transitive.

Assume therefore that we can find an essential set Δ with $\#(\Delta) = n - 2$ such that $G_{(\Delta)}$ is transitive on - $\setminus \Delta$. We will use the following relationship between orbits and stabilizers (e.g. Rotman [1996], p.123):

$$\#(G_\omega) = \frac{\#(G)}{\#\mathcal{O}(\omega)}, \forall \omega \in - \ .$$

Thus, if G is transitive on - , we have

$$\#(G_\omega) = \frac{\#(G)}{n}, \forall \omega \in - \ .$$

For $\omega' \in - \ \setminus \Delta$, let $\Delta' := \Delta \cup \{\omega'\}$ and $l = \#(\Delta')$. It follows by induction that

$$\#(G_{(\Delta')}) = \frac{\#(G)}{(n!/(n-l)!)},$$

and since $l = n - 1$, we have

$$\#(G_{(\Delta')}) = \frac{\#(G)}{n!}.$$

However, since Δ' contains $n - 1$ observations, $\#(G_{(\Delta')}) = 1$, so that it must be the case that $\#(G) = n!$. This is only possible if $G = S$, which contradicts $G < S$. □

Additional observations are likely to be essential in game settings where agents control the arrival of new observations through their actions. In that case, for a given language, our deterministic bound on the learning speed is appropriate. More generally, one might be interested in the consequences of a random selection of the language, and of observations being randomly chosen.[17] While detailed information about the general case is difficult to obtain, in can be noted that, as an immediate consequence of Proposition 5, whenever observations are drawn from a fixed distribution with full support, the expected time until a common language is learned is strictly less for G than for S , provided $G < S$.

6 Fast Learning with Absence of a Common Description

We are most interested in the case where an *a priori* absence of a common description is compatible with fast learning, both because this is the most

[17]Even in the case where observations are controlled by players actions, there may be randomness if choices are made simultaneously.

difficult case for fast learning, and because whenever there is no common language on - , there will be an absence of a common description on some nontrivial subset of - . So far we have shown that fast learning "is the rule" and that sometimes it can indeed occur in conjunction with an *a priori* absence of a common description. One may then be led to the conjecture that there is a rich set of environments (subgroups of S) with such a co-occurrence. This section shows that this conjecture can be confirmed in a qualified sense.

We will deal with the qualification first. Our first result in this section focusses on a particular class of co-occurrences of fast learning and *a priori* absence of a common description. These are learning patterns composed of a period of incremental learning followed by a jump to a full common description. Our result on this type of learning is essentially negative; there are definite restrictions on the nature of such co-occurrences. In particular, we will show that for sufficiently large n and most $l < n$, there does not exist a group $G \leq S$ such that a common language on - can be learned with l observations while $G_{(\Delta)}$ remains transitive on - $\setminus \Delta$ for all Δ with $\#\Delta < l$.

In order to prove this result we need to introduce a few additional concepts from the theory of permutation groups. Call any 2-cycle a *transposition.* One can show (e.g. Rotman [1996], p.63) that for $n \geq 2$ every $g \in S$ is a product of transpositions. If g can be factored into an even number of transpositions, then g is called an *even permutation.* The set of all even permutations in S forms a subgroup, A , that is referred to as the *alternating group* of degree n.

We also need the concept of a multiply transitive group. If G is a group acting on - , one can define an action on - ^{k} by

$$g(\omega_1, \ldots, \omega_k) := (g(\omega_1), \ldots, g(\omega_k)).$$

Consider - $^{(k)}$, that subset of - ^{k} that is composed of all those k-tuples whose elements are distinct; - $^{(k)}$ is G-invariant for all G and for all k. G is called *k-transitive* if G is transitive on - $^{(k)}$.[18]

The following facts about k-transitive groups will be useful (e.g. Dixon and Mortimer [1996], p.33, and Wielandt [1964], p.19). For $k > 1$, k-transitivity implies $(k-1)$-transitivity. G is k-transitive on - if and only if G_ω is $(k-1)$-transitive on - $\setminus \omega$. G is transitive if and only if it is 1-transitive. The alternating group A is (#- $- 2$)-transitive. Finally, in

[18]Multiply transitive groups made an early appearance in game theory in von Neumann and Morgenstern's [1947] discussion of symmetry in games.

addition to these elementary facts about multiply transitive groups, we will make use of the following result by Wielandt [1960][19]:

Theorem 1 *(Wielandt) Let $G \leq S$ be an 8-transitive group of finite degree. Then $G \geq A$.*

Wielandt's proof of this result assumes what is known as the Schreier Conjecture. The Schreier conjecture in turn can be established via the classification of finite simple groups. Actually, using this classification one can strengthen the result further to show that unless a finite permutation group contains the alternating group, it is at most 5-transitive (e.g. Dixon and Mortimer [1996], p.218).

Call an observation that identifies only the observed element itself an *incremental observation.* At the other extreme are observations that lead to a simultaneous identification of all the remaining elements of - ; those observations will be referred to as *revealing observations.* Note that when only two elements of - remain unidentified, then an additional observation is automatically revealing. Given S , for example, a common language is learned with $n - 2$ incremental observations, followed by one revealing observation. If g is an n-cycle, then given $< g >$ a common language is learned with zero incremental observations and a single revealing observation.

These two examples represent opposite ends of the spectrum of possible learning speeds. What about the intermediate ranges of the spectrum? Consider A_4, the alternating group of degree four. The set of rankings associated with A_4 is

$$A_4 = \begin{array}{cccccccccccc} 1 & 2 & 3 & 4 & 1 & 1 & 2 & 2 & 3 & 3 & 4 & 4 \\ 2 & 1 & 4 & 3 & 4 & 3 & 3 & 4 & 2 & 1 & 1 & 2 \\ 3 & 4 & 1 & 2 & 2 & 4 & 1 & 3 & 4 & 2 & 3 & 1 \\ 4 & 3 & 2 & 1 & 3 & 2 & 4 & 1 & 1 & 4 & 2 & 3 \end{array}$$

If the element corresponding to the second row is observed and the labeling rule assigns the label "1", this singles out the set of rankings

$$\begin{array}{ccc} 2 & 3 & 4 \\ 1 & 1 & 1 \\ 4 & 2 & 3 \\ 3 & 4 & 2 \end{array}$$

[19] A statement, proof and discussion of this result can also be found in Dixon and Mortimer [1996], p.218

Note that with the exception of the object that is now called "1", this collection of descriptions leaves every element unidentified. Formally, the stabilizer of the element "1" is transitive on the complement of "1;" the observation "1" is incremental. The properties of multiply transitive groups imply straightforwardly that instead of "1" we could have considered any other observation and would have obtained the same result. Thus, given that A_4 expresses the players' beliefs, the first observation is incremental. Note further, that in the induced subgroup only the identity fixes any of the remaining elements. Therefore, whatever the second observation, it will be revealing.

In summary, in the case where - has four elements, we can find permutation groups such that all observations are either incremental or revealing and where either the first observation is revealing ($< g >$, where g is a 4-cycle), the second observations is revealing (A_4), or the third observation is revealing (S_4). Conditional on all observations being revealing or incremental, a four-element set then permits the full range of possibilities. However, as indicated before, this example is the exception when it comes to exhausting the possibilities of combining incremental and revealing observations.

If G is a partial language on - that exhibits absence of a common description, and a common language can be learned with l incremental observations followed by one revealing one, we say that "- can be learned with l incremental observations." We saw that a four-element set could be learned with either 0, 1, or 2 incremental observations, depending on which partial language is in place; these are all the possibilities since the $(n-1)$st observations is always revealing. According to the next proposition this state of affairs is quite exceptional.

Proposition 6 *If* $\#- \geq 11$, *then a common language on - cannot be learned with* l *incremental observations whenever* $7 \leq l \leq \#- - 4$.

Proof. In order to arrive at a contradiction, suppose that $\#- \geq 11$ and that - can be learned with l incremental observations where l is in the specified interval. Then the group G that expresses the players' uncertainty over - must be $(l+1)$-transitive, i.e. still transitive on the complement of the observations after l observations, but not $(l+2)$-transitive for otherwise at least $l+1$ observations are needed to learn - . Thus, if $l \geq 7$, then G must be 8-transitive and therefore by Wielandt's theorem (Theorem 1), G contains the alternating group. But then G is $(\#- - 2)$-transitive and therefore the pointwise stabilizer of a set Δ with no more than $\#- - 3$ elements is still transitive on - $\setminus \Delta$, contrary to the assumption that the $l+1$st observation is revealing (which is part of the assumption that - can be learned with l incremental observations). □

Note that since Theorem 1 can be strengthened, the bounds on "learning with l incremental observations" could be tightened as well.[20]

Despite the last result, one can show that there is indeed a large set of scenarios in which an a priori absence of a common description is compatible with the possibility of fast learning. Of course, the previous result tells us that incremental observations do not play an important role in such learning. Most observations will be "partially revealing," i.e. besides identifying the observed object itself, they introduce identifying distinctions among the other objects that have not yet been observed.

Definition 3 *Given a partial language $G \leq S$,* a common language can be learned in k steps *if $G_{(\varpi)} \neq e, \quad \forall \varpi \in -^{k-1}$ and $\exists \hat{\omega} \in -^{k}$ such that $G_{(\hat{\omega})} = e$.*

In group theory, the set $\hat{\omega}$ is referred to as a "minimal basis" for G.

The following result shows that for a given size of the set of objects, - , even with absence of a common description, there are typically multiple partial languages that admit fast learning. In particular, it is shown that there exists a partial language that admits k-step learning whenever k is a divisor of #(-).

Proposition 7 *Let k be a divisor of #(-). Then there exists a partial language $G \leq S$, with absence of a common description, from which a common language can be learned in k steps.*

Proof. Let #(-) $= n$, and without loss of generality enumerate - as $1, \ldots, n$. Then there exists an integer l such that $n = k \times l$. For any subset K of S call the smallest subgroup of S containing K the *group generated by K*. Consider the group $H \leq S_n$ that is generated by the ("component") cycles

$$(il+1, \ldots, (i+1)l) \quad i = 0, \ldots, k-1$$

[20]Note also that our last result is of interest is because it illuminates the limits for modeling learning with a given *a priori* known structure. It is not difficult to construct games with rules that induce exactly the learning patterns ruled out in Proposition 6. Take for the example a game which involves picking identical balls from an urn, one per period. Each period a ball is picked, it is marked with the period number and returned to the urn. Except in the 51st period, where the rule prescribes that all remaining balls be marked with the numbers 51 through 100. Such phenomena arise if there are exogenous forces that alter the structure.

and by the cycle product

$$\prod_{j=1}^{l}(j, j+l, j+2l, \ldots, j+(k-1)l).$$

To verify transitivity, first note that if $\omega = 1$ belongs to an orbit, then all elements that are moved by the component cycle $(1, \ldots, l)$ belong to the same orbit. This follows from considering compositions of cycles. Similarly, examining powers of the cycle product, it follows that if 1 belongs to an orbit, $1+l,\ 1+2l, \ldots$ all belong to that orbit. Furthermore, all of these are moved by one of the different component cycles, and repeated application of those component cycles shows that for each such cycle all elements moved by that cycle belong to the orbit of $\omega = 1$. Thus all elements belong to the same orbit.

To see that k observations suffice to learn a common language from H, i.e. $\exists \hat{\omega}$ such that $H_{(\hat{\omega})} = e$, note that each observation ω removes those cycles from the set of remaining generators that move that element. The first observation therefore removes one component cycle *and* the cycle product. There are only $k-1$ cycles left that form the generators of G_ω. Then simply pick the remaining $k-1$ observations from different component cycles such that each of those observations eliminates one of those cycles from the set of remaining generators.

Finally, observe that $k-1$ (or fewer) observations are insufficient for learning a common language from H because there are k component cycles and each observation removes at most one component cycle from the set of generators. □

In interpreting this result, one should keep in mind that it is a statement about the possibilities for (or the variety of) fast learning, and not an assertion about the prevalence of fast learning. To fully address the prevalence question, one would have to make assumptions about the probability distribution over languages and use a much more detailed characterization of all the possible languages. This is beyond the scope of this paper. A few tentative comments can be made, however. First, with a uniform distribution over all languages, our earlier result (Proposition 5) implies that $n-1$-step learning is rare. Second, with a uniform distribution over partial languages that satisfy absence of a common description, our last result implies that $n-1$-step learning is again atypical. Note also that our last result is not exhaustive. For example, a look at the dihedral groups shows that for any number of objects, there is always a partial language that satisfies absence of a common description and admits 2-step learning.[21] Third, consider partial languages that already make some distinctions among individual objects, i.e. induce multiple orbits. Our results imply that each such orbit $\mathcal{O}$ can typically be learned with fewer than $\#(\mathcal{O}) - 1$ observations.

In light of our discussion in the introduction, our last result brings to mind multiple categorizations of objects as a source of structure in language. The construction in the proposition would correspond to a situation where we can sort objects into different categories, there is order both of and within the categories but there are no maximal (minimal) categories or maximal (minimal) elements within the categories. As mentioned earlier, categorization is also a source of structure when objects can be cross-categorized. As a practical matter the prevalence of fast learning phenomena or of structure in language in general may then, for example, be related to the number of categories that a set of players is aware of, and to their ability to commonly sort the categories.

7 Relation to the Literature

The problem of language learning, structure in language and creative use of language are relatively new in economics. Closest in spirit to the present paper is probably Rubinstein [1996] who is concerned with the structure of binary relations appearing in natural language. Like us, he has as one of his premises that "evolutionary forces make it more likely that the 'optimal' structures are observed [...]." He argues for certain properties that make binary relations in language more useful; among them the facility with which nameless elements in a set can be indicated, which reminds one of creative language use. He finds that this criterion of "indication-friendliness" is only satisfied by linear orderings.

For linguists of course, the questions of structure in language and how language is learned are central. Chomsky's research agenda for example attempts to identify a universal language faculty, a "generative grammar." Such a grammar would account for the fact that apparently relatively few observations suffice to learn a language that is capable of generating expressions of infinitely many meanings, in particular novel expressions that have not been encountered before. Thus viewed, the generative grammar accounts for creative language use (e.g. Chomsky [1988]).

[21] The dihedral groups correspond to the symmetries of regular polyhedra.

Within economics there is recent work by Segal [1996] that is related to ours. He starts with a common language but assumes that players cannot use the language to deduce a way of playing a coordination game. Instead they communicate with their common language within some organizational form. His objective is to determine the optimal organizational form (protocol, mechanism), where the likely quality of the communication outcome is traded off against a measure of communication complexity. Like in the present paper and in Rubinstein's work the focus is on characterizing efficient structures (protocols), and players are assumed to attain common knowledge of the structure via prior exposure to similar coordination problems. Segal shows that coordination by authority performs well in a wide range of circumstances and alludes to the possibility of using models of this kind to think of organizations as incomplete contracts.

In their study of optimal debate rules, Glazer and Rubinstein [1997] point out the potential role for natural binary relations (e.g. geographic proximity) in influencing persuasion rules. There may for example be an advantage in requiring that an argument that references a particular city be countered with reference to a city that belongs to the same natural category of cities. This is an example of the role of structure in an adversarial setting.

Wernerfelt [1998] studies how members of an organization construct languages in equilibrium. He postulates that agents have only access to a restricted set of codes. He finds that there may be benefits to using similar codes, and a centralized organization (with coarse code communication) may be preferred to decentralized decision making, even with costs for setting up the centralized organization.

Goyal and Janssen [1996] and Kramarz [1996] both reexamine the work of Crawford and Haller. Goyal and Janssen argue that optimal learning requires a prior convention to resolve multiplicities. Kramarz considers coordination with more than two players. He examines optimal learning, compares it to simple learning heuristics, and argues that learning rules that are highly history dependent may be inefficient because they may introduce more ambiguities than they resolve.

Two recent papers examine how coordination is affected or aided by structure. Chwe [1996] examines how structure affects collective action. In Chwe's model individuals share a common interest in coordinating collective action but have to rely on social networks to spread information about each individual's readiness to participate in collective action. The form of the network determines the speed at which information travels, the likelihood of collective action, and the time spent until collective action is achieved.

Calvert [1991] considers Crawford and Haller's "learning to coordinate" paradigm under constantly changing conditions. He declares that "[...] the basic problem of social order involves the achievement of coordination in fundamentally *new* situations [...]," and points out the benefit to players who make use of the "common-knowledge environment" of the games.

8 Conclusion

Communication and coordination in novel circumstances has to contend with language being at least partially incomplete. It is then important to know which constraints an incomplete language imposes on communication and coordination, and which possibilities it affords for learning a complete language. This paper isolates novelty of objects from common knowledge of structure and labeling rules. This permits one to develop a coherent framework in which to investigate the role of structure in coordination and learning. It is shown that structure aids both coordination and learning, even in the extreme case where there is no distinction among individual objects. In future work it would interesting to relax the common knowledge assumptions and to investigate frameworks that permit more general learning patterns (e.g. the ones ruled out by Proposition 6).

Appendix

A *group* G consists of a set, and an operation ,"$\circ$", on the set such that for all $f, g, h \in G$, (i) the associative law holds ($f \circ (g \circ h) = (f \circ g) \circ h$); (ii) there exists an identity element e such that $e \circ g = g = g \circ e$; and (iii) for all $g \in G$ there exists an inverse g^{-1} with $g \circ g^{-1} = e = g^{-1} \circ g$.

One group that is of particular interest to us is the set of *all* permutations of elements of the set - , denoted by S . Here the group operation is composition of permutations. This group is referred to as the *symmetric group* on - . If G is a group and H is a subset of G, then H is a *subgroup* of G, denoted $H \leq G$, if H contains the identity, and is closed under taking inverses and under the group operation.

Observation. *A set H of permutations satisfies the condition $Hg^{-1} = H\ \forall g \in H$ if and only if H is a subgroup of S .* **Proof.** Consider necessity first. (i) (existence of the identity). $Hg^{-1} = H$ and $g \in H$ together imply that $e = gg^{-1} \in H$. (ii) (existence of an inverse) $Hg^{-1} = H$ and $e \in H$ together imply that $g^{-1} = eg^{-1} \in H$. (iii) (composition of permutations in H defines an operation on H) ($Hg^{-1} = H, \forall g \in H$) and ($g^{-1} \in H, \forall g \in H$) implies that $Hg = H, \forall g \in H$ from which it follows that $fg \in H, \forall f, g \in H$. (iv) (associative law) associativity is inherited from S . Conversely, if H is a permutation group, then since H is closed under the group operation, we have $Hg \subset H, \forall g \in H$. Since $e \in H$, this implies that $Hg = H, \forall g \in H$, and in conjunction with $g^{-1} \in H, \forall g \in H$, we also have $Hg^{-1} = H, \forall g \in H$. □

Permutations can also be viewed as functions which leads to the notion of group G *acting* on a nonempty set - . For each $\omega \in$ - and for each $g \in G$, define an element $g(\omega) \in$ - . Then the group G *acts on* - if

(i) $e(\omega) = \omega \quad \forall \omega \in$ - , and

(ii) $h(g(\omega)) = hg(\omega) \quad \forall \omega \in$ - , and $\forall\ g, h \in G$.

If $g \in S$ and $\omega \in$ - , then we say that g *fixes* ω if $g(\omega) = \omega$; otherwise g *moves* ω. Let $\omega_1, \ldots, \omega_r \in$ - , $\omega_i \neq \omega_j$ for $1 \leq i, j \leq r$, $i \neq j$. If g fixes ω_j for $j \neq 1, \ldots, r$, and if $g(\omega_1) = \omega_2, g(\omega_2) = \omega_3, \ldots, g(\omega_{r-1}) = \omega_r, g(\omega_r) = \omega_1$, then g is called an *r-cycle*, and is denoted $(\omega_1, \omega_2, \ldots, \omega_r)$.

Let G be a group and $g \in G$. If $g^k = e$ for some $k \geq 1$, then the smallest such k is called the *order* of g. If G is a group and $g \in G$, then

$< g >:= \{g^n | n \in \mathcal{Z}\}$ is the *cyclic subgroup* of G that is *generated* by g. One convinces oneself easily that if g is an r-cycle, then g has order r and $< g >= \{e, g^1, \ldots, g^{r-1}\}$.

For any set $\Delta \subset -$, H a group acting on - , and $g \in G$ let

$$\Delta^g := \{g(\omega) \in - \, | \omega \in \Delta\}.$$

Then one can define the *setwise stabilizer* of the set Δ as

$$H_{\{\Delta\}} = \{g \in H | \Delta^g = \Delta\},$$

and the *pointwise stabilizer* of the same set as

$$G_{(\Delta)} = \{g \in G | \omega^g = \omega, \forall \omega \in \Delta\}.$$

One checks easily that for any Δ, both pointwise and setwise stabilizers are subgroups of H. The stabilizer of a point ω will be denoted by G_ω.

References

Aitchison, J. [1993], *Linguistics,* Chicago: NTC Publishing Group.

Aumann, R. and A. Brandenburger [1995], "Epistemic Conditions for Nash Equilibrium," *Econometrica,* **63**, 1161-1180.

Bacharach, M. [1993], "Variable Universe Games," in K. Binmore, A. Kirman, and P. Tani (eds.), *Frontiers of Game Theory,* Cambridge, MA: MIT Press.

Bacharach, M. and M. Bernasconi [1997] "The Variable Frame Theory of Focal Points: An Experimental Study," *Games and Economic Behavior,* **19**, 1-45.

Batali, J. [1996], "Computational Simulations of the Emergence of Grammar," University of California-San Diego, Department of Cognitive Science Working Paper.

Blume, A. and Gneezy, U. [1998], "An Experimental Investigation of Optimal Learning in Games," University of Iowa and University of Haifa, Department of Economics Working Paper.

Blume, A., D.V. DeJong, Y.-G. Kim and G.B. Sprinkle [1998], "Experimental Evidence on the Evolution of the Meaning of Messages in Sender-Receiver Games," forthcoming in *American Economic Review.*

Blume, A. [1996], "Neighborhood Stability in Sender-Receiver Games," *Games and Economic Behavior,* **13**, 2-25.

Blume, A., Y.-G. Kim and J. Sobel [1993], "Evolutionary Stability in Games of Communication," *Games and Economic Behavior,* **5**, 547-575.

Calvert, R.L. [1991], "Elements of a Theory of Society among Rational Actors," University of Rochester, Department of Political Science Working Paper.

Chomsky, N. [1990], "Language and the Problem of Knowledge," in E.P. Martinich (ed.) [1996], *The Philosophy of Language,* New York: Oxford University Press.

Chwe, M. S.-Y. [1996], "Structure and Strategy in Collective Action: Communication and Coordination in Social Networks," University of Chicago, Department of Economics Working Paper.

Crawford, V. and H. Haller [1990], "Learning how to Cooperate: Optimal Play in Repeated Coordination Games," *Econometrica,* **58**, 581-596.

Dixon, J.D. and B. Mortimer [1996], *Permutation Groups,* New York: Springer Verlag.

Fuller, J.G. [1962], *The Gentleman Conspirators: The Story of the Price-Fixers in the Electrical Industry,* New York: Grove Press.

Glazer, J. and A. Rubinstein [1997], *Debates and Decisions: On a Rationale of Argumentation Rules,* Tel Aviv University, Working Paper.

Goyal, S. and M. Janssen, [1996], "Can We Rationally Learn to Coordinate," *Theory and Decision,* **40**, 29-49.

Kramarz, F. [1996], "Dynamical Focal Points in N-Person Coordination Games," *Theory and Decision,* **40**, 277-313.

Rotman, J. [1996] *A First Course in Abstract Algebra,* Upper Saddle River, New Jersey: Prentice Hall.

Rubinstein, A. [1996] "Why are Certain Properties of Binary Relations Relatively More Common in Natural Language?" *Econometrica,* **64**, 343-355.

Schelling, T.C. [1960], *The Strategy of Conflict,* Cambridge: Harvard University Press.

Segal, Ilya [1996] *Communication Complexity and Coordination by Authority,* University of California-Berkeley, Department of Economics Working Paper.

von Neumann, J. and O. Morgenstern [1947], *Theory of Games and Economic Behaviour,* Princeton: Princeton University Press.

Wärneryd, K. [1993], "Cheap Talk, Coordination, and Evolutionary Stability," *Games and Economic Behavior,* **5**, 532-546.

WERNERFELT, B. [1998], "Organizational Languages," MIT, Department of Management Science Working Paper.

WIELANDT, H. [1960] "Über den Transitivitätsgrad von Permutationsgruppen," *Math. Z.* **74**, 297-298.

WIELANDT, H. [1964] *Finite Permutation Groups*, New York: Academic Press.