



Testing for the emergence of spontaneous order

Konstantinos Georgalos¹ · John Hey²

Received: 7 December 2018 / Revised: 12 November 2019 / Accepted: 27 November 2019 /
Published online: 10 December 2019
© The Author(s) 2019

Abstract

We report on an experimental investigation of the emergence of Spontaneous Order, the idea that societies can co-ordinate, without government intervention, on a form of society that is good for its citizens, as described by Adam Smith. Our experimental design is based on a production game with a convex input provision possibility frontier, where subjects have to choose a point on this frontier. We start with a simple society consisting of just two people, two inputs, one final good and in which the production process exhibits returns to specialisation. We then study more complex societies by increasing the size of the society (groups of 6 and 9 subjects) and the number of inputs (6 and 9 inputs respectively), as well as the combinations of inputs that each subject can provide. This form of production can be characterised as a cooperative game, where the Nash equilibrium predicts that the optimal outcome is achieved when each member of this society specialises in the provision of a single input. Based on this framework, we investigate whether Spontaneous Order can emerge, without it being imposed by the government. We find strong evidence in favour of the emergence of Spontaneous Order, with communication being an important factor. Using text classification algorithms (Multinomial Naive Bayes) we quantitatively analyse the available chat data and we provide insight into the kind of communication that fosters specialisation in the absence of external involvement. We note that, while communication has been shown to foster coordination in other contexts (for example, in public goods games, market entry games and competitive coordination games) this contribution is in the context of a production game where specialisation is crucial.

Keywords Spontaneous order · Co-ordination · Production · Specialisation · Communication

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10683-019-09637-8>) contains supplementary material, which is available to authorized users.

✉ Konstantinos Georgalos
k.georgalos@lancaster.ac.uk

Extended author information available on the last page of the article

Springer

1 Introduction

The idea of Spontaneous Order—that societies can co-ordinate without government intervention—is mainly attributed to Adam Smith, but some think that it goes back further, even as far back as the fourth century BC, and the Chinese philosopher Zhuang Zhou who argued that “good order results spontaneously when things are let alone”. The idea was further developed by the French philosopher (and anarchist) Proudhon in the nineteenth century, and played a major role in the thinking of the Scottish Enlightenment, being immortalised in Adam Smith’s *Invisible Hand*. Smith developed the concept of the division of labour and argued that rational self-interest and competition can lead to economic prosperity. Michael Polanyi (1948) was the first to actually call this process Spontaneous Order, a notion that the Austrian School of Economics would later refine and make it the flagship of its social and economic thought, mainly expressed by Karl Menger, Ludwig von Mises, and Friedrich Hayek, with Menger wondering “How can it be that institutions which serve the common welfare and are extremely significant for its development come into being without a common will directed toward establishing them?” (Menger 1985, p.146). More recently, Sugden (1989) provided a thorough exposition of the importance of Spontaneous Order for studying economics, by discussing how many of the institutions in the market economy are conventions that no one has designed but they have simply evolved. Sugden concludes by stating “Thus the study of spontaneous order may help to explain why we have some of the moral beliefs that we do have, without in any way being able to show that we ought to have them”.

We experimentally investigate whether and how Spontaneous Order emerges, building on Smith’s idea of the division of labour—as demonstrated in his Pin Factory example. In this, he showed that in a production process, if different workers specialise in different parts of the production process, then the workers can jointly produce a much greater volume of output than if they do not: when all workers specialise we have an efficient outcome. Moreover, he argued that this would happen ‘spontaneously’ and without external intervention.

We investigate this hypothesis, first with just 2 workers, then with 6 and finally with 9—to explore whether the number of workers influences the speed of convergence to the efficient outcome. We also examine the crucial influence of *communication* on convergence. By this, we mean communication between the workers, and not outside involvement.

This paper starts with a review of relevant background experimental research, providing the motivation for this study. We then describe the basic model, and then, in Sect. 4 we discuss extensions to, and variations on, this basic model. The experimental implementation is described in Sect. 5. Section 6 presents a description of behaviour in the experiment, while Sect. 7 analyses the text messages between subjects and sheds light on why they did what they did. Section 8 concludes.

2 Background material

Central to our concern is the idea of Spontaneous Order—that societies can spontaneously co-ordinate in such a way as to maximise social welfare—without government intervention. The co-ordination required is in two fields: production and exchange. In the former, members of society should co-ordinate in such a way that production is achieved in a Pareto efficient way; and hence so that output cannot be increased by changing the production process. In the industrial organisation literature, this process is also known as *Entrepreneurial Discovery*, with the Austrian school advocating that “a chief virtue of the free enterprise system is its evocation of the discovery (and fulfilment) of opportunities for social betterment”, (Hayek 1978). Similarly, Kirzner (1985) was arguing that under a regime of economic freedom, substantial and socially beneficial epiphanies occur more often. In the latter, exchange should also be carried out in a way that is Pareto efficient; and hence, so that welfare cannot be increased by changing the exchange process. Hayek (1945) theorised that markets implement competitive equilibrium prices and allocations when information of the economic agents is decentralized and private, a notion that became later known as the *Hayek Hypothesis* in the experimental markets literature (Smith 1982). Our focus is on the emergence of Spontaneous Order in Entrepreneurial Discovery.

There have been a large number of experimental investigations of the efficiency of the *exchange* process (the Hayek Hypothesis), initiated by the seminal work of Smith (1962, 1982). Most of these have considered only exchange and not production. The basic design is simple and clever: there is a hypothetical good being traded. Some of the subjects are designated as potential buyers of the good, and are each told their reservation values for one or more units of the good. The other subjects are designated as potential sellers of the good, and are each told their reservation values for one or more units of the good. Trade takes place through a double auction method. Buyers are rewarded with the difference between their reservation value and the price that they pay. Sellers are rewarded with the difference between the price that they receive and their reservation value. The overwhelming message is that markets are generally competitive and hence efficient, though efficiency does decline as the markets become more complicated—for example when they are dynamic as in Smith et al. (1988) (a useful survey of the dynamics of such exchange experiments can be found in Crockett 2013). An interesting extension which seems to incorporate production is that of Shachat and Zhang (2015), in which buyers are incentivised in the same way—by giving each a set of reservation values—but the sellers’ reservation values (costs) are choosable (under constraints) by the sellers. In a sense, they can choose the cost of production and its spread over time. This, however, is not coordination with others of production.

On the other hand, studies on Entrepreneurial Discovery, Hayek’s (1978) conjecture that the free enterprise system is the most effective in making discoveries, are very few. Demmert and Klein (2003) is the first study that aims to test the Hayek/Kirzner conjecture. In their experiment, they tested the influence of

various motivation and incentive schemes on entrepreneurial discovery. Due to a very narrow definition of entrepreneurial discovery, along with some experimental design constraints, the authors do not manage to find support in favour of the Hayek/Kirzner conjecture. On top of that, their study does not include any social interaction element, neither has it included specialisation.¹

A key exception that studies both production and exchange is Crockett et al. (2009), which is the main stimulus for the experiment reported in this paper. In Crockett et al. (2009), subjects were exposed to a decision problem in which they not only had to decide on their production, but also on how they would exchange their output with other subjects in the experiment. In order to reach an efficient situation, subjects had to learn about their own production possibilities, and also had to learn about how to trade and with whom, so as to take advantage of their specialised skills. Subjects could exploit their competitive advantage in the production of certain goods, but only if they exchanged the produced goods in an efficient way. As the authors admit, the experiment was quite a daunting one for the subjects, and therefore it is not surprising that convergence to equilibrium was slower than in a purely exchange environment. Moreover, they state “It has been over 230 years since Adam Smith articulated the proposition that specialisation creates wealth and that specialisation is in turn supported by exchange; yet after all this time we have no theory of the discovery process that supports exchange and specialisation, nor an understanding of what impedes it”. In order to understand which part of the Crockett et al. experiment caused the slower convergence, we ran an experiment with just production and not exchange. There are important differences between our setup and that of Crockett et al. In the latter, subjects could directly choose a combination of final (consumption) goods to produce; in ours, subjects could choose what inputs to provide into a production function. In Crockett et al., there were two types of subjects, with different types having different production functions, and with each type of subject having a comparative advantage in the production of one final good. Production possibility frontiers for different types differed; they were approximately linear with differing slopes for the different types.

In our setup, as we wanted an experiment as close as possible to Adam Smith’s pin factory, in which specialisation led to efficiencies, we gave each subject an *input possibility frontier*, and asked them to choose what inputs they wanted to provide. There was a single output (‘pins’ to complete the analogy) and the volume of output was determined by the aggregate quantities of the various inputs, aggregated over the subjects. We can think of the inputs as particular skills. In the *Wealth of Nations*, Smith writes of the different operations required to produce a pin:

drawing out the wire; straightening it; cutting it; pointing it; grinding it at the top for receiving the head; making the head (which requires two or three distinct operations); putting it on; whitening the pins; putting them into the paper.

¹ The game theoretic literature has explored some of the notions mentioned above, such as the Spontaneous Order in market entry games (Erev and Rapoport 1998) or the epiphany in finding the optimal play in games (Dufwenberg et al. 2010) among others. To the best of our knowledge, none includes coordination of the subjects in order to discover the optimal strategy in the spirit of entrepreneurial discovery.

In all, he counted about eighteen distinct operations, or inputs. He noted that if each worker worked in isolation, “they certainly could not each of them have made twenty, perhaps not one pin in a day”, whereas, if they specialised, “ten persons ... could make among them upwards of forty-eight thousand pins in a day”. Specialisation could produce enormous gains.

While in the real world, different people have different innate skills, we simplified the experiment by making them all *ex ante* identical—by giving them identical *input provision possibility frontiers*. So our story is as follows (we give details below): each subject decides independently what quantities of the various inputs to provide (subject to the constraint of the input provision possibility frontier); the inputs are aggregated and output is produced with the aggregate inputs. The output is distributed amongst the subjects depending upon their individual contributions (we give details below). The aggregate production function was such that if the aggregate amount provided of any one input was zero, the output would be zero. This implies the necessity for co-ordination. People should specialise, but given the identical *input provision possibility frontiers*, it was not clear in what they should specialise. This is a marked difference from Crockett et al. in which the production possibility frontiers gave a clear signal. We now give detail.

3 The basic model

We start simple, with just ($m=$) 2 workers and ($n=$) 2 inputs into a production process with a single output. Each worker can decide how much of two inputs he or she will provide. Obviously there is a constraint on this decision, and we represent this by the following concave *input provision possibility frontier*, where x_{ij} denotes the number of units of input j chosen by worker i ($i = 1, \dots, m, j = 1, \dots, n$):

$$\sum_{j=1}^m x_{ij}^d = A^d \quad \text{where} \quad 0 < d < 1 \quad (\text{for } i = 1, \dots, n) \tag{1}$$

For $A = 100$ $d = 0.5$ and $m = 2$ this gives the Fig. 1 below. Each worker can choose any point on this curve. For example, they could decide to provide 100 units of input 1 and none of input 2, or equal quantities (25) of each, or none of input 1 and 100 of input 2. Any point on this curve can be chosen.

Having determined their individual provisions of the two inputs, these are then aggregated into total provisions of the two inputs, $X_i = \sum_{j=1}^m x_{ij}$ for $i = 1 \dots, n$, which are then used to produce a single output through the Cobb–Douglas function

$$V = \prod_{i=1}^n X_i^{1/n} \tag{2}$$

We note crucially that if the *total* provision of either input is zero then the value of output is zero.

Output having been determined, we now specify how the workers are paid. First the value of output, V , is divided equally between the two inputs (as they are

Fig. 1 An input provision possibility frontier

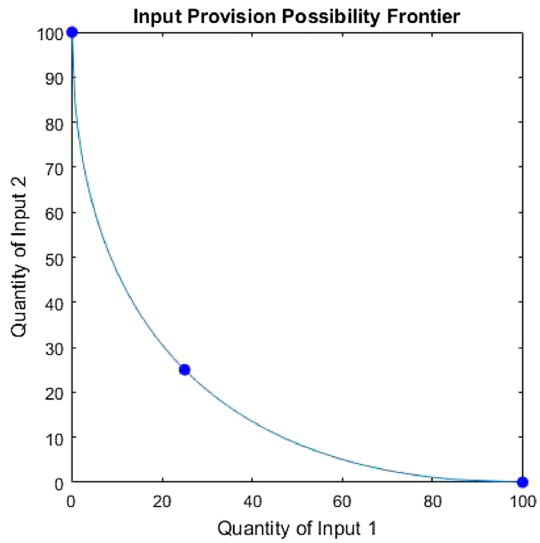


Table 1 The sequence of decisions and consequences

	Input 1	Input 2	Payoff from Input 1	Payoff from Input 2	Total Payoff
Worker 1	31	19	20.1	7.3	27.4
Worker 2	9	48	5.8	28.6	24.4
Total	40	67	25.9	25.9	51.8

Table 2 Consequences if both workers specialise in the provision of one input

	Input 1	Input 2	Payoff from Input 1	Payoff from Input 2	Total payoff
Worker 1	100	0	50	0	50
Worker 2	0	100	0	50	50
Total	100	100	50	50	100

symmetric in its production), and then allocated to the two workers on the basis of the fraction of each input that each provided. We give an example below. The arrows indicate the sequence of events (Table 1).

The workers choose their input provisions (those in bold): in this example, worker 1: 31 of input 1 and 9 of input 2; worker 2: 9 of input 1 and 48 of input 2. These imply total provisions of 40 of input 1 and 67 of input 2. These produce output valued at 51.8. This is divided equally between the two inputs—giving 25.9 to each. Then, because worker 1 provided 31 of the 40 units of input 1 that were provided, he or she gets a fraction 31/40 of 25.9, which is 20.1, for his or her provision of input 1; because worker 1 provided 19 of the 67 units of input 2 that were provided, he

Table 3 Consequences if both workers provide equal quantities of the two inputs

	Input 1	Input 2	Payoff from Input 1	Payoff from Input 2	Total payoff
Worker 1	25	25	12.5	12.5	25
Worker 2	25	25	12.5	12.5	25
Total	50	50	25	25	50

or she gets a fraction $19/67$ of 25.9, which is 7.3 for his or her provision of input 2. The total pay for worker 1 is thus $20.1 + 7.3 = 27.4$. The same method is applied for worker 2.

If, instead, the workers each specialised in one input then we would get the table below (Table 2):

They would both earn considerably more. Indeed it can be shown that perfect specialisation leads to the highest payoffs for both, and that this is a Nash equilibrium.

If, on the contrary they each decided to provide equal quantities of the two inputs—as in the table below—they would both be considerably worse off (Table 3).

However, perfect specialisation is not the obvious thing to do—particularly as the value of output is zero if there is zero provision of either input: so if both specialised in the provision of the same input they would both earn nothing. Some element of coordination is necessary. This was made possible—but not inevitable—by the facility of *communication* in the experiment. We allowed communication by having a message box through which subjects could send messages to other subjects. We saved all the messages sent by subjects.

4 Extensions to, and variants on, the basic model

The basic story has just two workers. We extended it to 6 workers and to 9 workers. In each of these extensions, workers had to choose *three* inputs, constrained by the obvious generalisation of Eq. (1). Total output was determined by Eq. (2), with $n=6$ or 9 as appropriate. For both 6 and 9 inputs we had two variants, described by the ‘overlappingness’ of the set of three inputs produced by the workers. We will give detail later. For now, the crucial point is the same: payments are maximised if all specialise in one of the three inputs, but there must be co-ordination between the workers: once again, if two workers decide to specialise in the provision of the same input they will all get a zero payment.

5 Treatments and the experimental implementation

We ran 24 sessions, each with either 12 or 18 subjects. We vary the size of the groups, the type of communication, as well as the combination of inputs that each subject could provide. Below we provide the rationale behind the various treatments. The common hypothesis for all the treatments, is that providing subjects with

the ability to communicate freely is sufficient to establish specialization in input provision, and therefore to allow groups to discover the optimal combination of inputs and maximise welfare for the society. This argument leads to Hypothesis 1.

Hypothesis 1 Spontaneous order can evolve without conscious human design and can maintain itself without there being any formal machinery for enforcing it.

Since the early 60 s, the effect of group size on the level of cooperation has been a topic of considerable controversy with three main sides of the argument. First, there are theories which argue that cooperation decreases as the number of participants in a group increases (Olson 1965; Hardin 1982). Then, there is theoretical work suggesting that the level of cooperation might increase with the size of the group (Chamberlin 1974; McGuire 1974). Finally, there is the idea that there exists an inverted U relationship between coordination and the size of a group, according to which medium-size groups tend to cooperate more than smaller groups, while large groups tend to cooperate less than medium-size groups (Poteete and Ostrom 2004). The experimental research on the topic has not managed to provide enlightening evidence in favour of any of the theories, with some studies finding a negative effect (Grujic et al. 2012; Nosenzo et al. 2015), with others finding a positive effect (Isaac et al. 1994; Zelmer 2003) or a curvilinear effect (Capraro and Barcelo 2015).

Nevertheless, most of the empirical research is based on experiments on Public Good provision, where the free-rider problem may affect players' willingness to cooperate. While in our context there is absence of free-riding incentives, increasing the size of the group (and consequently the number of inputs) may lead to lower levels of cooperation due to the increased level of complexity or to the increased probability of one of the group members making a mistake. We ran sessions with groups of size 2, 6 or 9 according to the treatment. This argument leads to Hypothesis 2.

Hypothesis 2 Increasing the size of the groups will have a negative impact on the coordination and therefore the level of specialisation.

For the 6 and 9 sessions, we had overlapping and non-overlapping treatments. 'Non-overlapping' means that subjects 1, 2 and 3 each provided inputs 1, 2 and 3; that subjects 4, 5 and 6 each provided inputs 4, 5 and 6; and, in the 9's, that subjects 7, 8 and 9 each provided inputs 7, 8 and 9. 'Overlapping' means that subject 1 provided inputs 1, 2 and 3; that subject 2 provided inputs 2, 3 and 4; that subject 3 provided inputs 3, 4 and 5; and so on. Our hypothesis is that in the non-overlapping treatments it would be easier for subjects to co-ordinate their decisions.

The Schelling's hypothesis (Schelling 1960) is a solution concept from game theory which predicts that in the absence of communication, players will choose those strategies that seem natural or salient due to some property that all the players can recognise (focal point). Moreover, the *salience* bias (Kahneman et al. 1982) refers to the tendency of the individuals to focus on items that are more prominent or emotionally striking. Experimental evidence has shown that

subjects tend to adopt focal points (Mehta et al. 1994; Isoni et al. 2013) as well as to exhibit salience bias (Tiefenbeck et al. 2018).

Observing the emergence of Spontaneous Order in our experiment could be the result of subjects choosing a salient option. For example, in the case of two workers and two inputs, one could argue that the two salient choices are either to provide equal quantities of the two inputs, or for either of the workers to specialise only in one of the inputs. The same argument can be extended for larger sizes of groups and inputs. In a 6-worker, 6-input environment, where inputs 1, 2 and 3 are provided only by three subjects (non-overlapping), it may be more salient, and easier to figure out, that each member of the group should specialise in one of the three inputs. In the case of overlapping inputs, inputs 1, 2 and 3 are now provided by 5 different subjects, which may make coordination much more difficult. We agree that the salient decisions (specialise in one input, and produce equal quantities of the three inputs) are the same, but the coordination problems are much more intense in the overlapping treatments: in these treatments, a subject has to try to coordinate with 5 other subjects (each of whom are trying to coordinate with 5 others, and so on); in the non-overlapping treatments, subjects have to coordinate with just 2 others. Moreover, visually, in the experiment, when only three subjects provide the three inputs, the results are presented in a 3×3 matrix, while in the non-overlapping treatments they are spread over a 9×9 matrix making it more difficult to identify that specialisation may lead to higher payoffs. This argument leads to Hypothesis 3.

Hypothesis 3 In non-overlapping treatments it would be easier for subjects to coordinate their decisions compared to the overlapping treatments.

The fact that communication usually increases coordination and speeds up convergence to optimality is something that has been empirically supported in various occasions (see Crawford 1998; Crawford 2019 for a review). In all the treatments presented above, the communication is ‘Full’, meaning that all messages would be sent to and from all members in the group. Nevertheless, Schelling’s intention was to apply his theory in *tacit* bargaining situations, where communication is incomplete or even impossible. In the remaining treatments, we vary the kind of communication facilities available to subjects to test whether it would have a significant effect to the levels of cooperation. In the ‘None’ communication treatment, no messaging was allowed. Thus, the lack of communication may cause inefficiency. This argument leads to Hypothesis 4.

Hypothesis 4 Coordination will be harder to be achieved without communication.

We then manipulate the kind of communication available to the subjects. Using a minimum-effort coordination game, Weber (2006) shows that, even though efficient coordination does not occur in groups that start off large, efficiently coordinated large groups can be “grown”, by starting with small groups that find it easier to coordinate and then add entrants—who are aware of the group’s history. Essentially Weber is saying that coordination is easier in small groups than in large. Our

'*between-inputs*' communication treatment splits the whole group into sub-groups. In this, only the subjects providing the same inputs could send messages to each other.² Compared with the 'full information' treatment, subjects were sending messages to, and receiving messages from, a smaller number of other subjects; in a sense subjects are in 'small groups' in the 'between-inputs' communication treatment, and in large groups in the 'full information' treatment, so one might expect better coordination in the former. This argument leads to Hypothesis 5.

Hypothesis 5 Coordination in the between-inputs communication treatments will be easier compared to treatments where full communication is allowed.

A main question in the economics of organization literature is whether organizations are more efficient when they are centrally managed or when decentralization takes place and the various units within an organization function independently. This topic has been the subject of several recent experimental studies [see Evdokimov and Garfagnini 2019; Hamman and Martínez-Carrasco 2018; Brandts and Cooper 2018; Cooper et al. (1989)] with the evidence being inconclusive and heavily dependent on the context. In our '*Leader*' treatment, one subject was chosen (randomly by the software) to be the leader, and he or she could send messages to all the other members of the group, and they could send messages to him or her; but no other messages were allowed. Cooper et al. (1989) show results, which if carried over to our context, would suggest that the leader, by directing everyone to a specialised input, may actually work better than full communication. We expect that coordination will be easier compared to the full communication one. This argument leads to Hypothesis 6.

Hypothesis 6 Introducing a leader to coordinate the production process will make coordination easier compared to the treatments where full communication is allowed.

The experiment was carried out in the EXEC laboratory at the University of York, with subjects recruited using *hroot*. Written Instructions (please see the online appendix) were placed on the subjects' desks and these were read aloud by one of the experimenters over the tannoy system. Any questions were publicly answered; there were few.

In the 6's sessions we recruited either 12 or 18 participants. They were randomly divided into two or three 6-person groups; they never knew the identities of the other members of their group. In the 9's sessions we recruited 18 participants. They were randomly divided into two 9-person groups; once again, they never knew the identities of the other members of their group.

In order to get sufficient data for analysis, and to give the subjects a chance to learn about the problem, we gave the subjects three identical problems and

² The implementation of this differed in the 'overlapping' sessions and the 'non-overlapping' treatments.

Table 4 Number of independent observations by treatment and communication

Treatment ↓ Communication →	FULL	LEADER	BETWEEN	NONE
2	1*8*3=24	0	0	0
6/NON-OVERLAP	4*3*3=36	0	0	0
6/OVERLAP	4*3*3=36	1*2*3=6	0	0
9/NON-OVERLAP	4*2*3=36	1*2*3=6	1*2*3=6	1*2*3=6
9/OVERLAP	4*2*3=36	1*2*3=6	1*2*3=6	1*2*3=6

Key to entries: number of sessions × number of groups × number of problems

We denote the treatments as N/OVERLAPPINGNESS/COMMUNICATION. N is the size of the group and takes the values 2, 6 or 9. OVERLAPPINGNESS is the kind of combination of inputs and can be either overlapping (OVERLAP) or non-overlapping (NON-OVERLAP). COMMUNICATION is the kind of allowed communication in a treatment and takes the values FULL, LEADER, BETWEEN and NONE

seven repetitions of each. The groups changed composition between problems. Inevitably, there was some overlapping between members of the groups, across problems. We take this into consideration in our subsequent analysis. Moreover to make the experiment fair, we paid them their payoff on the *final repetition* of a randomly-chosen problem. They were clearly told that the first six repetitions of each problem did *not* count towards payment, and so could be used for practice and persuasion.

To summarise, in all of our treatments, we vary the size of the groups, the combination of inputs that each subject could provide, as well as the type of communication. We had treatments with groups of three different sizes: 2, 6 and 9-member groups. The combination of inputs could be either overlapping or non-overlapping (when the size of the group was greater than two). Finally, the communication could be full, between leader and the rest of the group, between group members providing the same inputs or no communication at all. In what follows, we denote each of the treatments as N/OVERLAPPINGNESS/COMMUNICATION, with N indicating the size of the group, OVERLAPPINGNESS the kind of combination of inputs and COMMUNICATION the type of allowed communication. For instance, the treatment 6/OVERLAP/FULL, stands for the treatment with 6-member groups, overlapping provision of inputs and full communication between all the members of the group. The details of the different treatments are in Table 4.

6 The experimental findings

We display visually the results of the sessions in Fig. 2. The lines are the means of the sum of the inputs provided by each of the subjects. If all the subjects managed to coordinate correctly this sum would be 100; if each tried to equalise their provision of the three inputs, the sum would be 33 1/3. On the horizontal axis is the repetition. Recall that subjects were given 7 repetitions of each of 3 problems; between

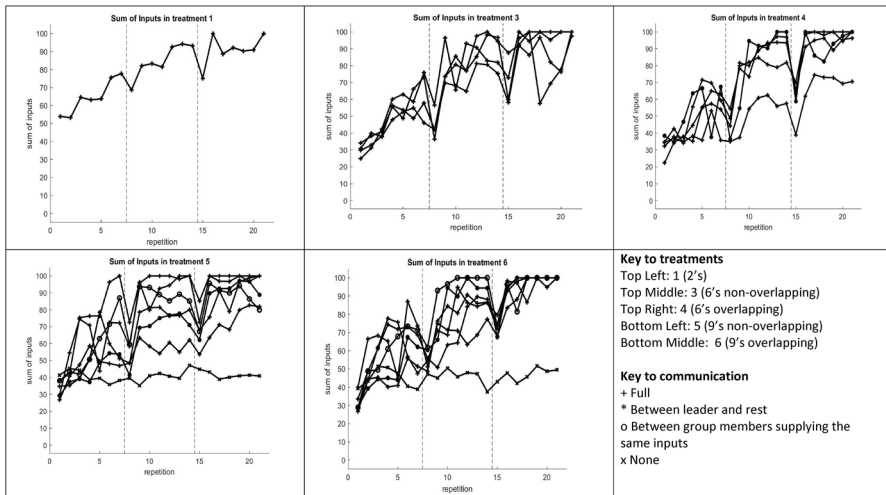


Fig. 2 The aggregate supply of inputs over the various treatments

problems the groups were changed and they were randomly re-matched with new partners. The graph plots all three problems consecutively, but note that repetitions 8 and 15 were with new group members—indicated by the dashed vertical lines.

One immediate finding is that communication is essential. Examine the two lines with no communication—the bottom lines in the graphs on the lower rows in Fig. 2. (A detailed analysis indicates that there are some subjects who have worked out what is best to do, and try and signal that) but these signals are not picked up by the others. So without communication we do not get spontaneous order (Hypothesis 4).

With communication things are different. With full communication in the 2/ FULL treatment between both members of the group, we get complete specialisation by the final repetition, with the final two subjects being drawn into line on that final repetition. A similar thing happens in all the other treatments irrespective of the form of communication. Visually it is difficult to see much difference in behaviour of the subjects with the different kinds of communication. Similarly, there do not seem to be any treatment effects (Hypothesis 1).

We can investigate deeper with regression analyses.³ We take as the dependent variable that graphed in Fig. 2: the mean of the sum of the inputs provided by each of the subjects. One obvious explanatory variable is the repetition number (counted overall in the session⁴). To take account of the fact that group members changed

³ We have confirmed that our results are significant based on non-parametric analysis (Wilcoxon rank sum test). The treatment effects are in line with those reported in the manuscript. Nevertheless, we decided to report treatment effects based on the regressions as this provides us with an additional quantitative measure of how important the difference between treatments is.

⁴ Repetitions 1–7 are the 7 repetitions of the first problem; 8–14 the 7 repetitions of the second problem; 15–21 the 7 repetitions of the third problem.

Table 5 Communication effects

	Estimate	SE	t-Stat	p value
<i>(Intercept)</i>	18.61	2.26	8.24	0.000
<i>t</i>	2.453	0.10	25.71	0.000
<i>frp</i>	-15.601	1.65	-9.45	0.000
<i>full</i>	29.376	2.07	14.18	0.000
<i>leader</i>	28.614	2.53	11.31	0.000
<i>between</i>	35.164	2.77	12.68	0.000

Adjusted R-Squared 0.686

t repetition number. *frp* first repetition dummy. *full* takes the value 1 if full communication; 0 otherwise. *leader* takes the value 1 if communication between the leader and the group members 0; otherwise. *between* takes the value 1 if communication between the subjects providing the same inputs; 0 otherwise. No communication is the default

between problems, we also include a ‘first repetition dummy’ *frp*, which takes the value 1 in the first repetition of a problem and 0 elsewhere.

Let us look first at the communication effects. To do this we introduce three communication dummies: *full*, *leader* and *between*, which take the value 1 in the corresponding communication treatment, and 0 elsewhere. The results are in Table 5. As the default is no communication, the estimated coefficients tell us how much each of the other three communication treatments affected the relationship between the mean sum of inputs and the repetition. Full communication does better than communication only between the leader and the other members of the group (a result which goes against our Hypothesis 6). But interestingly full communication does not have the strongest effect: that between the subjects providing the same inputs does better (Hypothesis 5). This suggests that closer messaging contacts are better for co-ordination: each member does not need to persuade all the group as to what to do, only those providing the same inputs as he or she. Obviously, in the overlapping treatments, communication will have a ripple effect. All the coefficients are positive and significant confirming that with the lack of communication it is not possible for Spontaneous Order to emerge (Hypothesis 4).

We now turn to the treatment effects. We introduce treatment dummies, denoted by *two*, *6non-overlap*, *6overlap*, and *9non-overlap*, which take the value 1 in the corresponding communication treatment. We also use these interacted with the repetition variable. The results are in Table 6. It will be seen that not all coefficients are significant, but we can interpret their implications. The default is the 9/OVERLAP treatment and it will be seen that *repetition* has a significant positive effect. Indeed in this 9/OVERLAP treatment the mean sum of inputs starts at 50.11 ($47.73 + 2.38$) in the first repetition and ends at 97.71 ($47.73 + 21 \times 2.38$) in the final period of the final problem. Treatments 6/NON-OVERLAP and 6/OVERLAP start off slightly

Table 6 Treatment effects

	Estimate	SE	t-Stat	p value
(Intercept)	47.73	2.61	18.28	0.000
<i>t</i>	2.38	0.20	11.61	0.000
<i>frp</i>	-15.60	1.94	-8.02	0.000
<i>two</i>	14.18	7.24	1.96	0.051
<i>6non-overlap</i>	-6.41	4.24	-1.51	0.131
<i>6overlap</i>	-8.04	3.96	-2.03	0.043
<i>9non-overlap</i>	1.18	3.62	0.33	0.743
<i>two</i> × <i>t</i>	-0.51	0.58	-0.89	0.376
<i>6non-overlap</i> × <i>t</i>	0.56	0.34	1.66	0.097
<i>6overlap</i> × <i>t</i>	0.52	0.32	1.66	0.098
<i>9non-overlap</i> × <i>t</i>	-0.37	0.29	-1.27	0.204

Adjusted R-Squared 0.565

t: repetition number. *frp*: first repetition dummy. *two*: takes the value 1 if 2's treatment 0; otherwise. *6non-overlap*: takes the value 1 if 6's non-overlapping treatment; 0 otherwise. *6overlap*: takes the value 1 if 6's overlapping treatment; 0 otherwise. *9non-overlap*: takes the value 1 if 9's non-overlapping treatment; 0 otherwise. *treatment* × *t*: is the interaction of the variable *treatment* with the repetition number. 9/OVERLAP treatment is the default

lower but end up slightly higher. Treatments 2/FULL and 9/NON-OVERLAP do the opposite: starting off slightly higher but ending slightly lower. Regarding the size of the groups, it appears that small groups (2's) and large groups (9's) are the most efficient, while medium groups (6's) do worse, indicating a U-shape relationship between size and coordination (Hypothesis 2).

Regarding the effect of 'overlappingness', while the sign of the coefficients is in the predicted direction (9/NON-OVERLAP and 9/NON-OVERLAP treatments do better than the 6/OVERLAP AND 9/OVERLAP respectively), the result does not seem to be significant (Hypothesis 3).

We note that in both these regressions, the *first repetition dummy* is significantly negative, indicating the effect of changing group composition on behaviour: subjects do not trust new people.

Finally we report in Table 7 the effects of both the treatment and the communication, but here omitting interaction terms. Once again communication, and the first period dummy, have the most effect, with the same pattern emerging. The treatment effects are less strong both in terms of significance and of sign.

Table 7 All effects

	Estimate	SE	t-Stat	p Value
(Intercept)	20.03	2.34	8.58	0.000
<i>t</i>	2.45	0.09	26.17	0.000
<i>frd</i>	-15.60	1.62	-9.62	0.000
<i>two</i>	4.73	2.97	1.59	0.111
<i>6non-overlap</i>	-4.06	1.80	-2.25	0.025
<i>6overlap</i>	-6.01	1.66	-3.62	0.000
<i>9non-overlap</i>	-2.84	1.46	-1.95	0.051
<i>full</i>	30.72	2.14	14.36	0.000
<i>leader</i>	30.14	2.54	11.89	0.000
<i>between</i>	35.16	2.72	12.91	0.000

Adjusted R-Squared 0.697

t: repetition number. *frd*: first repetition dummy. *6non-overlap*: takes the value 1 if 6's non-overlapping treatment; 0 otherwise. *6overlap*: takes the value 1 if 6's overlapping treatment; 0 otherwise. *9non-overlap*: takes the value 1 if 9's non-overlapping treatment; 0 otherwise. *full*: takes the value 1 if full communication; 0 otherwise. *leader*: takes the value 1 if communication between the leader and the group members 0; otherwise. *between*: takes the value 1 if communication between the subjects providing the same inputs; 0 otherwise. 9/OVERLAPPING is the default

7 Communication content analysis

In this section we study the effect that communication has on the decision-making of the participants, and particularly on their choice of total input provision. In the literature, there have been suggested various different methods of analysing communication data from experiments. The different approaches that one can follow in order to analyse a chat content can be classified as:

- Content analysis (widely used in social psychology) where third party coders are involved in classification of messages into predetermined categories as in Cooper and Kagel (2005), Goeree and Yariv (2011), Sutter and Strassmair (2009), Bougheas et al. (2013) or Chen and Houser (2017).
- Self-classification of messages, which is similar to the content analysis but now the authors generate the labels and also classify data themselves, as in Charness and Dufwenberg (2006) or Schotter and Sopher (2007).
- Descriptive analysis, where the authors use a particular extract from the data in order to make a point, without using any quantitative measures, as in Crockett et al. (2009), Kimbrough et al. (2008).
- Quantitative analysis based on the text rather than human coders, where the text is mined for keywords as in Zhang and Casari (2012), Penczynski (2016), Moellers et al. (2017).

Table 8 Relative frequency of classified messages

Category	t1c1	t3c1	t4c1	t5c1	t6c1	t5c2	t6c2	t5c3	t6c3
Specialise	0.16	0.18	0.18	0.17	0.22	0.20	0.28	0.17	0.26
Confusion	0.04	0.02	0.06	0.03	0.11	0.04	0.05	0.03	0.09
Proposal	0.41	0.37	0.38	0.37	0.40	0.44	0.47	0.36	0.41
Coordination	0.46	0.56	0.53	0.56	0.56	0.40	0.54	0.56	0.59
History	0.23	0.29	0.27	0.32	0.33	0.30	0.32	0.26	0.30
Total	422	1407	1484	1532	1463	200	196	317	302

Note As a message can be classified in more than just one category, the frequencies do not add up to 1

We have raw data of 16,510 messages from all treatments with communication (536 pages of text). Due to the large volume of data, we resort to text mining methods in order to classify the data. More particularly, we apply a commonly used machine learning algorithm, the *Multinomial Naïve Bayes Classifier*. The first task in text-mining is to tokenise each message (break message into words) to create a term-document matrix. In this matrix, each row represents one document (in our case a message) and each column represents one term (token). Each entry of the matrix contains the frequency of that term in that document. To reduce the size of the matrix we use the common practices of removing all *stopwords* such as “the”, “we”, “are”, “to”, “a”, “for” and reducing the words to their stem (for example have and having are reduced to “hav”).⁵ We also disregard all sentences with three or fewer words, given that very rarely they provided a message worth classifying. This leaves a sample of 8277 messages to classify. After reading a representative sample of the data from all the treatments, we established the following 5 categories:

- Specialise: a message suggesting to maximise the input of only one of the inputs.
- Confusion: a statement that the subject is confused regarding the task or the strategy.
- Proposal: a message proposing a strategy or seeking for advice on what to do.
- Coordination: a message referring to coordination, trust and generally creating a team spirit.
- History: a message that makes reference to the previous rounds or problems.

Using 1200 messages as a training sample for the classifier, we classified all messages. Table 8 reports the relative frequency of classified messages for each of the treatments. Note that a message could be classified in more than one categories. For all the treatments, messages that propose (or seek for advice) along with messages that promote coordination are the most frequent. Messages expressing confusion were the less frequent.

⁵ The natural language toolkit NLTK for Python was used to clean the text and apply the classification algorithm. The Lancaster stemmer was used to stem the words.

In order to investigate the effects that each category of communication has on the total provision of inputs, we follow Sutter and Strassmair (2009) and estimate linear models with random effects on the subject level using Feasible Generalised Least Squares (FGLS). We pool the data of each treatment along with those of the no communication treatments (there are 756 observations without communication) and we use dummies for the 5 communication categories along with the repetition number. The dependent variable is the mean of the sum of the inputs provided by each of the subjects. As the default is no communication, the estimated coefficients tell us how much each of the five communication categories affected the relationship between the mean sum of inputs and the repetition.

To save on space, the estimation Tables are delegated to the online appendix. A shared finding between all treatments is that the coefficients for the categories of messages that promote a team spirit (Coordination), messages where subjects seek or provide suggestions on what to do (Proposal) and messages referring to the past actions of the group (History), are always significant and have a positive sign. Out of these three categories, messages of Proposal and Coordination have, on average, the strongest effect. It is interesting that the Proposal effect is much stronger in the 9-input cases, indicating that the more complex that task becomes, the greater the need for advice. Messages on past actions (History) have a positive impact on the average input, since such messages usually refer to successful combinations in previous repetitions, or strategies that worked in previous problems with other groups, sharing in this way their past experience. Regarding the effect of specialisation, although it has the expected positive sign, it is not significant in all treatments and it does not seem to have the greatest impact on the choices of the subjects, as it might be expected. The coefficient for messages that suggest Confusion is positive and significant in some of the treatments. This could be the case where subjects clearly expressed their confusion, and exchange of messages allowed them to identify the best plan of action.

8 Conclusions

In this paper we report on an experiment designed to investigate whether and how Spontaneous Order emerges in a production environment that supports specialisation, but where there is a lack of a suitable institution to co-ordinate the subjects to achieve the socially optimal outcome. Our experimental design is based on a production process with a single output, where each worker has to decide how much of various inputs to provide. The social welfare is maximised when all the subjects specialise in one input, in an environment where specialisation is not the obvious thing to do. The subjects must first discover that specialisation generates the highest outcome, and then must find a way to co-ordinate in order to decide the way specialisation should take place. We vary the number of workers, the number of inputs, the overlappingness of inputs, as well as the way subjects could communicate (or not) within a group, in an effort to investigate the impact of these variables on the speed of convergence to the efficient outcome.

Our results can be summarised as follows:

1. Communication is crucial for the emergence of the Spontaneous order.
2. In all variants with communication, Spontaneous Order emerged and most subjects opted for complete specialisation by the final repetition.
3. Full communication does better than communication between the leader and the members of the group.
4. Communication between the same providers of inputs does better compared with all other forms of communication.
5. Repetition has a significantly positive effect.
6. Subjects do not trust other subjects when groups are re-matched.
7. Messages that carry a cooperative spirit, and messages that seek or give advice, have the highest impact towards specialisation.

It should not come as a surprise that communication is a crucial factor in achieving coordination. Adam Smith argued that language provides the foundation that makes coordination and cooperation possible.⁶ As Levy (1997) highlights “Adam Smith, especially, argues that being human is the same as using language. Reason and speech are primitives for him; we no more choose to use language than we choose to be human. His argument in *Wealth of Nations* is that trade and language are two aspects of the same process; humans trade because we have language, non-humans do not trade because they do not.”

We note that, while communication has been shown to foster coordination in other contexts (for example, in public goods games, market entry games and competitive coordination games), our contribution is in the context of a production game where specialisation is crucial but not obvious. In these other contexts, communication usually seems to lead to the socially optimal outcome (as distinct from the Nash Equilibrium). In our experiment, coordination implies reaching a socially optimal outcome, which *is* a Nash Equilibrium. In our experiment there are several Nash Equilibria, all equally good from a social perspective, and the social objective is to coordinate on one of these. We show that this usually happens: order is usually attained—spontaneously and without external intervention.

Recent literature has provided evidence that Spontaneous Order emerges in exchange. Crockett et al. (2009) in an experiment with both production and exchange, found that Spontaneous Order emerges for some of the participants, but the benefits from specialisation could be enjoyed, conditional on exchanging the produced goods in an efficient way, that was also to be discovered. Our objective in this experiment was to study the emergence of Spontaneous Order by disentangling production from exchange, in an effort to identify whether the production part of the experiment impeded the convergence to equilibrium. Using a simple experimental design, we obtain a clear result that Spontaneous Order in production can emerge in the absence of a central authority, but under the condition that communication takes place. One may therefore speculate that the reason for the low convergence in Crockett et al. (2009) could have been that subjects needed to coordinate in both

⁶ See Adam Smith (1761). We are grateful to a careful anonymous referee for pointing out this reference to us.

production and the exchange process. The latter can have an important impact on the decisions of a firm or a government regarding the organisational structure of a particular entity.

Potential extensions of our experimental design could include both production and exchange but with different subjects getting involved in one or the other. It could include the involvement of managers, who manage the workers but do not provide any tangible input to the production process; and the emergence of managers, where subjects can evolve from workers to managers; or the production of more than one good, potentially with different required levels of specialisation for each product. We leave these questions for future work.

Acknowledgements The authors would like to thank the Editors of the journal, Lata Gangadharan and Roberto Weber, and an anonymous referee for very helpful comments that led to significant improvements in both the analysis and the presentation. The financial aid of the British Academy is gratefully recognised (Grant SG 153403).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bougheas, S., Nieboer, S., & Sefton, M. (2013). Risk-taking in social settings: Group and peer effects. *Journal of Economic Behavior & Organization*, 92, 273–283.
- Brandts, J., & Cooper, D. (2018). “Truth be told: An experimental study of communication and centralization. Working Paper 1046, Barcelona Graduate School of Economics.
- Capraro, V., & Barcelo, H. (2015). Group size effect on cooperation in one-shot social dilemmas II: Curvilinear effect. *PLoS ONE*, 10, e0131419.
- Chamberlin, J. (1974). Provision of collective goods as a function of group size. *American Political Science Review*, 68, 707–716.
- Charness, G., & Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, 74, 1579–1601.
- Chen, J., & Houser, D. (2017). Promises and lies: Can observers detect deception in written messages. *Experimental Economics*, 20(2), 396–419.
- Cooper, R., De Jong, D. V., Forsythe, R., & Ross, T. W. (1989). Communication in the battle of the sexes game. *RAND*, 20, 568–587.
- Cooper, D., & Kagel, J. H. (2005). Are two heads better than one? Team versus individual play in signaling games. *American Economic Review*, 95, 477–509.
- Crawford, V. (1998). A survey of experiments on communication via cheap talk. *Journal of Economic Theory*, 78(2), 286–298.
- Crawford, V. (2019). Experiments on cognition, communication, coordination, and cooperation in relationships. *Annual Review of Economics*, 11, 167–191.
- Crockett, S. (2013). Price dynamics in general equilibrium experiments. *Journal of Economic Surveys*, 27, 421–438.
- Crockett, S., Smith, V. L., & Wilson, B. J. (2009). Exchange and specialisation as a discovery process. *Economic Journal*, 119, 1162–1188.
- Demmert, H., & Klein, D. B. (2003). Experiment on entrepreneurial discovery: An attempt to demonstrate the conjecture of Hayek and Kirzner. *Journal of Economic Behaviour and Organization*, 50, 295–310.

- Dufwenberg, M., Sundaram, R., & Butler, D. (2010). Epiphany in the Game of 21. *Journal of Economic Behavior & Organization*, 75, 132–143.
- Erev, I., & Rapoport, A. (1998). Coordination, “magic”, and reinforcement learning in a market entry game. *Games and Economic Behavior*, 23(2), 146–175.
- Evdokimov, P., & Garfagnini, U. (2019). “Communication and behavior in organizations: An experiment. *Quantitative Economics*, 10, 775–801.
- Goeree, J. K., & Yariv, I. (2011). An experimental study of collective deliberation. *Econometrica*, 79, 893–921.
- Grujić, J., Eke, B., Cabrales, A., Cuesta, J. A., & Sánchez, A. (2012). Three is a crowd in iterated prisoner’s dilemmas: experimental evidence on reciprocal behavior. *Scientific Reports*, 2, 638.
- Hamman, J., & Martínez-Carrasco, M. A. (2018). Delegation and team selection in organizations: An experimental study. Working Paper.
- Hardin, R. (1982). *Collective action*. Baltimore: Johns Hopkins University Press.
- Hayek, F. A. (1945). The use of knowledge in society. *The American Economic Review*, 35(4), 519–530.
- Hayek, F.A. (1978). Competition as a discovery procedure. In: *New studies in philosophy, politics, economics and the history of ideas*. University of Chicago Press, Chicago, pp. 179–190.
- Isaac, M. R., Walker, J. M., & Williams, A. W. (1994). Group size and the voluntary provision of public goods. *Journal of Public Economics*, 54, 1–36.
- Isoni, A., Poulsen, A., Sugden, R., & Tsutsui, K. (2013). « Focal points in tacit bargaining problems: Experimental evidence. *European Economic Review*, 59, 167–188.
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Kimbrough, E., Smith, L. S., & Wilson, B. J. (2008). Historical property rights, sociality, and the emergence of impersonal exchange in long-distance trade. *American Economic Review*, 98(3), 1009–1039.
- Kirzner, I. M. (1985). *Discovery and the capitalist process*. Chicago: University of Chicago Press.
- Levy, D. (1997). Adam Smith’s rational choice linguistics. *Economic Inquiry*, 35(3), 672–678.
- McGuire, M. C. (1974). “Group size, group homogeneity, and the aggregate provision of a pure public good under Cournot behaviour. *Public Choice*, 18, 107–126.
- Mehta, J., Starmer, C., & Sugden, R. (1994). Focal points in pure coordination games: An experimental investigation. *Theory and Decision*, 36(2), 163–185.
- Menger, C. (1985). In L. Schneider & F. Nock (Ed.) *Investigations into the method of the social sciences with special reference to economics*. New York: NYU Press.
- Moellers, C. H., Normann, C., & Snyder, M. (2017). Communication in vertical markets: Experimental evidence. *International Journal of Industrial Organization*, 50, 214–258.
- Nosenzo, D., Quercia, S., & Sefton, M. (2015). Cooperation in small groups: The effects of group size. *Experimental Economics*, 18, 4–14.
- Olson, M. (1965). *The logic of collective action: Public goods and the theory of groups*. Cambridge: Harvard University Press.
- Penczynski, S. (2019). Using machine learning for communication classification. *Experimental Economics*, 22(4), 1002–1029.
- Polanyi, M. (1948). Planning and spontaneous order. *The Manchester School*, 16(3), 237–268.
- Poteete, A. R., & Ostrom, E. (2004). Heterogeneity, group size and collective action: The role of institutions in forest management. *Development and Change*, 35, 435–461.
- Schelling, T. C. (1960). *The strategy of conflict* (1st edn). Cambridge: Harvard University Press.
- Schotter, A., & Sopher, B. (2007). Advice and behavior in intergenerational ultimatum games: An experimental approach. *Games and Economic Behavior*, 58, 365–393.
- Shachat, J., & Zhang, Z. (2015). The Hayek hypothesis and long-run competitive equilibrium: An experimental investigation. *Economic Journal*, 127, 199–228.
- Smith, A. (1761). “Considerations concerning the first formation of languages, and the different genius of original and compounded languages”. *The Philological Miscellany I*, pp. 440–79. London.
- Smith, V. L. (1962). An experimental study of competitive market behavior. *Journal of Political Economy*, 70, 111–137.
- Smith, V. L. (1982). Markets as economizers of information: Experimental examination of the ‘Hayek hypothesis’. *Economic Inquiry*, 20, 165–179.
- Smith, V. L., Suchanek, G. L., & Williams, A. W. (1988). Bubbles, crashes, and endogenous expectations in experimental spot asset markets. *Econometrica*, 56, 1119–1151.
- Sugden, R. (1989). Spontaneous order. *The Journal of Economic Perspectives*, 3(4), 85–97.

- Sutter, M., & Strassmair, C. (2009). Communication, cooperation and collusion in team tournaments—An experimental study. *Games and Economic Behavior*, *66*, 506–525.
- Tiefenbeck, V., Goette, L., Degen, K., Tasic, V., Fleisch, E., Lalive, R., et al. (2018). Overcoming salience bias: How real-time feedback fosters resource conservation. *Management Science*, *64*(3), 1458–1476.
- Weber, R. (2006). Managing growth to achieve efficient coordination in large groups. *The American Economic Review*, *96*(1), 114–126.
- Zelmer, J. (2003). Linear public goods experiments: A meta-analysis. *Experimental Economics*, *6*, 299–310.
- Zhang, J., & Casari, M. (2012). How groups reach agreements in risky choices. *Economic Inquiry*, *50*, 502–515.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Konstantinos Georgalos¹  · John Hey²

John Hey
john.hey@york.ac.uk

- ¹ Department of Economics, Lancaster University Management School, Lancaster LA1 4YX, U.K.
- ² Department of Economics, University of York, Heslington, York YO10 5DD, U.K.