Title
Self-Organized Division of Cognitive Labor

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Authors
Andrade-Lotero, Edgar
Goldstone, Robert L.

Publication Date
2019

Peer reviewed

Self-Organized Division of Cognitive Labor

Edgar Andrade-Lotero (edgar.andrade@urosario.edu.co)
Department of Applied Mathematics and Computer Science, Carrera 6 No. 12C-06, Of. 506
Bogotá, COLOMBIA

Robert L. Goldstone (rgoldsto@indiana.edu)
Department of Psychological and Brain Sciences, Psychology Building 338
Bloomington, IN 53706 USA

Abstract

The division of labor phenomenon has been observed with respect to both manual and cognitive labor, but there is no clear understanding of the intra- and inter-individual mechanisms that allow for its emergence, especially when there are multiple divisions possible and communication is limited. Situations fitting this description include individuals in a group splitting a geographical region for resource harvesting without explicit negotiation, or a couple tacitly negotiating the hour of the day for each to shower so that there is sufficient hot water. We studied this phenomenon by means of an iterative two-person game where multiple divisions are possible, but no explicit communication is allowed. Our results suggest that there are a limited number of biases toward divisions of labor, which serve as attractors in the dynamics of dyadic coordination. However, unlike Schelling's focal points, these biases do not attract players' attention at the onset of the interaction, but are only revealed and consolidated by the in-game dynamics of dyadic interaction.

Keywords: Group cognition; Divergent behavioral norms; Focal points; Cooperation.

Introduction

An individual can often benefit from participating in a group when (s)he can perform just one component of the group's task while other individuals take care of other parts. When the other individuals also benefit from this arrangement, we speak of an efficient division of labor. For example, two roommates can choose between (a) preparing their lunch for themselves every day, and (b) dividing the days of the week on which one prepares lunch for two. In the latter case, both roommates benefit from not having to cook every day.

The benefits of division of labor have been studied not only with respect to manual labor (Smith, 2008), but also with respect to cognitive labor (Sloman & Fernbach, 2017; Kennedy, Eberhart, & Shi, 2001). For instance, one study showed that the puzzle of assigning categories to the nodes of a network such that no adjacent nodes have the same category could be efficiently solved as a self-organized, collective task if each individual is assigned to a single node and is only concerned about the acceptability of their local sub-network (Kearns, Suri, & Montfort, 2006).

In some collective groups, such as ant colonies or beehives, the division of labor occurs as a genetically designed organization (Weitekamp, Libbrecht, & Keller, 2017; Robinson, 1992). However, it can also emerge as a self-organized process, without leaders or explicit negotiations (Heylighen, 2013). For example, when a group of individuals has to collectively guess a target number, where the collective guess is the sum of their individual guesses, and the only feedback they receive is for how much their collective guess is greater or lesser than the target, individuals spontaneously differentiate their behaviors to either react or not react to the feedback, and the extent to which role differentiation occurs is predictive of group performance (Roberts & Goldstone, 2011).

What are the cognitive mechanisms that facilitate the self-organized division of labor? One possibility is that it arises from the principle of maximization of expected utility. In our previous example of the two roommates, successful division of the days of the week might be said to arise because it constitutes a Nash equilibrium, that is, a combination of choices in which no roommate can obtain a higher payoff by changing only their choice—fixing the other roommate's choice (Ross, 2018). However, as it turns out, maximization of expected utility is not sufficient to explain why roommates act in accord with a particular Nash equilibrium instead of another (Arthur, 1994; Colman, 2003). Some scholars have suggested that games with multiple Nash equilibria are not solved on the basis of maximization of expected utility, but rather by means of rough-and-ready rules of thumb based on limited knowledge and time. This approach is known as 'bounded rationality' to emphasize that people frequently have memory, attentional, and calculation limitations that prevent them from employing perfectly rational strategies (Holbrook, 2002; Simon, 1957). It could be claimed, returning to our roommates example, that the division of labor according to which Roommate A prepares lunch only on weekdays and Roommate B only prepares lunch on the weekend is achieved because they cannot think of a different division, or because this division is the most natural for both of them, even though there are many other possible divisions. This is an example of the focal point approach, according to which the set of all possible Nash equilibria is reduced to just a single point that is psychologically salient for all players (Mehta, Starmer, & Sugden, 1994; Schelling, 1960). Another possible proposal is that individuals possess a small set of simple strategies that they can apply in their search for a division. For example, they may stick to one strategy for as long as it provides acceptable results, and when it fails, they would swap it for another in their strategy set (a win-stay, lose-shift heuristic). There are cooperative scenarios, such as the famous El Farol problem

(Arthur, 1994), in which this heuristic works well. Another strategy could be to adapt one's own reactivity to the task based on how much the whole group is contributing. Indeed, as pointed out by Roberts and Goldstone (2011), there may be situations in which an individual helps the collective effort by refraining from acting or reducing their activity, allowing the other players to dominate the task.

We studied this phenomenon by means of an iterative two-person game where multiple divisions are possible, but no explicit communication is allowed. Our results suggest that there are a limited number of biases toward divisions of labor and that they work as attractors in the negotiation dynamics. Unlike Schelling's focal points, these biases do not attract players' attention at the onset of the interaction, but are only revealed and consolidated by the in-game dynamics of the dyadic interaction. In other words, these biases do not determine players' *a priori* actions, via some sort of iterative reasoning, for dividing up their task. Rather, the attractors only become salient as a result of the interaction.

Materials and methods

Participants and procedure

Participants were 90 undergraduate students at Indiana University in Bloomington who received course credits for approximately 1 hour of participation. Participants were run in 10 experimental sessions, each one requiring an even number of participants to be grouped into dyads. If an odd number of participants turned up to the session, one of them was randomly chosen and sent home. The number of dyads in each session were as follows: 4, 5, 3, 6, 4, 2, 6, 3, 8, and 4. Participants sat in a university computer lab, each at a sound- and sight-isolated personal computer running a version of the game implemented in the nodeGame platform (Balietti, 2017). The computer randomly paired participants into dyads and each dyad participated in 60 rounds of the game. Participants were instructed not to talk to each other and were not informed about who was paired with whom.

The task

The task is a two-player game, which we dub "Seeking the unicorn," in which players interact with 64 tiles arranged in an 8×8 grid (see the top panel in Figure 1). The grid can either hide a unicorn beneath one of the tiles or else the unicorn can be absent from the grid, either event can occur with equal probability. At the beginning of each round, the computer chooses with equal probability whether or not there is a unicorn, and if there is one, it randomly chooses a tile in which to hide it, each tile having an equal probability of being chosen. Then, players seek for the unicorn by uncovering tiles one at a time, with both players uncovering tiles simultaneously, in order to see what lies beneath them. What tiles have been uncovered and whether there is or not a unicorn is only known to the player that uncovers these tiles. Tiles uncovered by both players instantly change their color and both players can see this. At any time during the round, each player can guess whether the unicorn is present or absent. The other player will know this player's decision and can use it to inform their own guess. The round ends when both players announce that their guess is a final decision, and then they are shown their scores (see the bottom panel in Figure 1). The score depends on whether the player's guess is correct (32 points) or incorrect (-64 points), subtracting the number of tiles that were uncovered by both players.

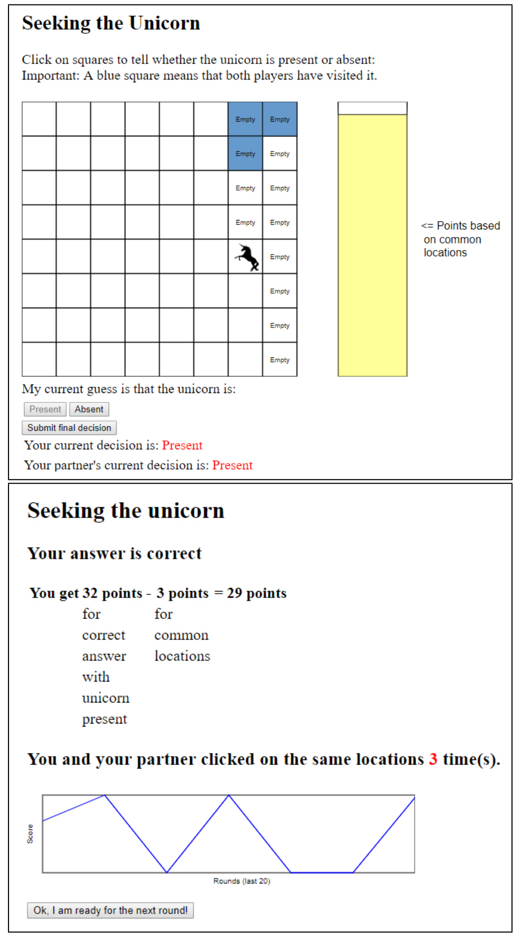

Figure 1: The experimental task. The top panel shows the grid as displayed to each player. By uncovering a tile, they know whether it is empty or contains the unicorn. Such information is private for the player. Tiles uncovered by both players have a blue background and both can see this coloration. They also have access to each other's guesses. The yellow column on the right decreases as the number of overlapping tiles increases. The round ends when both players submit their final decision. In the bottom panel we show the screen displaying the score and the score history over the last 20 rounds.

Measures

The following measure, which we call the Division of Labor Index (DLINDEX), determines the extent to which players split the grid into complementary regions:

$$\text{DLINDEX} = \frac{\begin{array}{c}\text{Tiles uncovered by one}\\ \text{or both of the players}\end{array} - \begin{array}{c}\text{Overlapping}\\ \text{tiles}\end{array}}{\text{Tiles in the grid}}$$

This measure instantiates the intuition that it is beneficial if a dyad collectively uncovers all of the tiles (first term) and does not overlap in any tiles uncovered (second term). Observe that it ranges from 0 to 1 with 1 being ideal division of labor and 0 being least efficient. There is only one way of being ideal, namely, when both players uncover the entire grid and do not overlap at all. Additionally, we measure how consistently a player uncovers tiles from one round to the next:

$$\text{Consistency}_n = \frac{\begin{array}{c}\text{Overlapping uncovered tiles from}\\ \text{Round } n-1 \text{ to Round } n\end{array}}{\begin{array}{c}\text{Tiles uncovered in}\\ \text{either of the two rounds}\end{array}}$$

This number ranges from 0 to 1 with 1 meaning that the player uncovers the same tiles on both rounds, and 0 meaning that the player uncovers a completely different set of tiles from one round to the next. We also define the distance and the similarity between two regions a and b in the grid in the following way:

$$\text{dist}(a,b) = \sqrt{\sum_{t \in \text{Tiles}} (a_t - b_t)^2}, \quad \text{sim}(a,b,\varepsilon) = e^{-\varepsilon * \text{dist}(a,b)}$$

Here, t represents the t-th tile in the 8×8 grid (represented as the list $[1, \ldots, 64]$), and a_t and b_t can be either 1 or 0, representing whether or not tile t belongs to a and b, respectively. The parameter ε in the definition of sim determines the extent to which the distance between two regions determines the similarity between them and, unless explicitly stated otherwise, we assume that ε=1.

Results

We should note at the outset that we have not used the entire dataset in our analysis. The reason is that rounds on which the unicorn is present provide us only with partial information as to how players split the grid, because on those rounds players do not have to uncover every tile. Once they find the unicorn, they will say that the unicorn is present and finish the round. Therefore, unless explicitly stated otherwise, we are only reporting results for trials on which the unicorn is absent.

For each dyad we created a figure displaying two grids, one for each player. In this figure we magnitude-coded each tile according to the number of times the player selected it through 60 rounds of the experiment in such a way that the darker the tile, the more times it was selected. Figure 2 shows the types of regions that were obtained. There were only four stable, successful pairs of complementary regions in the grid: the Left-Right, Top-Bottom, All-Nothing, and Inside-Outside splits. We call them the focal splits. Only dyads in the focal splits obtained an above-average DLINDEX, except for one dyad with no discernible stable region that nevertheless has an average DLINDEX of 0.82 (this is over 0.45 standard deviations above the average of the 45 dyads; this dyad determined the Mix type of split in Figure 2). We conclude that 26 out of 45 dyads successfully split the grid. This represents over 57% success in self-organizing division of labor.

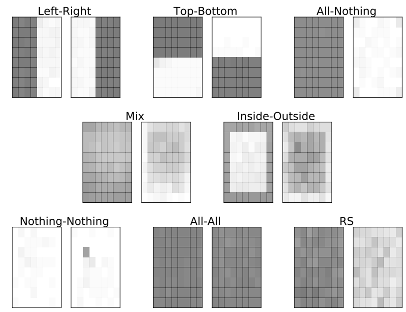

Figure 2: The seven types of splits of the grid that could be observed from our data. Each panel shows two grids, one for each player, with the regions uncovered through 60 rounds. The darker the tile, the more rounds the player uncovered it.

If our paradigm were a one-shot task in which players had to converge on a split of the grid on only one round, our data show that the average DLINDEX would be close to 0.43 (s.d.$\approx$ 0.32). By comparison, in our iterated task, the average DLINDEX rose up to almost 0.68 after 60 rounds (s.d.$\approx$ 0.32). The difference between these averages is statistically significant ($p < 0.001$). This shows that an efficient division of labor does not emerge on the first round, and that the iterated nature of our task facilitates its emergence.

But how did the division of labor emerge? We observed that, in general, dyads moved from lower to higher levels of DLINDEX, and that players in a low-level dyad tended to more frequently change their tile selection strategy from one round to the next with respect to players in a high-level dyad. Moreover, we found a positive correlation between a player's consistency on Round n and their score on Round $n-1$ ($\beta \approx 0.51$; $p < 0.001$). This supports the hypothesis that players used, at least to some extent, a win-stay, lose-shift heuristic (WSLS). That is to say, if their score is relatively high, which often occurs when the dyad splits the grid into complementary regions, each player tends to re-select their previously selected tiles; but if their score is low, they will be more likely to shift to different tiles.

However, WSLS does not seem to account for all the characteristics of the dyadic interaction. When we predict DLINDEX as a function of consistency, we see that, perhaps not surprisingly, dyads consisting of individuals who are relatively consistent in their tile selection strategies tend to divide labor better ($\beta \approx 0.36$; $p < 0.001$). However, we also observe an interaction such that dyads with players that differ in their consistencies tend to divide labor better than predicted when players have a large amount of overlap in their selected tiles. That is, if both players overlap considerably, it is best if one player is consistent and the other player is not. The evidence for this claim comes from comparing the linear regression model above with a model that includes the interaction between, on the one hand, the absolute difference in consistency between players on a given round and, on the other hand, the number of overlapping tiles on the previous round:

$$\text{DL}\textsc{index}(n) \sim \alpha + \beta_1 * \text{Consistency}(n) + \beta_2 * \text{difConsist}(n) + \beta_3 * \text{Overlap}(n-1) + \beta_4 * \text{difConsist}(n) * \text{Overlap}(n-1)$$

Our data show that this interaction is positive ($\beta_4 \approx 0.01$; $p < 0.001$). Moreover, an analysis of variance test ($p < 0.001$) confirms that this interaction effect accounts for significantly more variance in performance relative to the main effects. These results indicate that dyads eventually tend to most effectively divide labor despite initially overlapping in their tiles when one player is consistent/stubborn and the other player is inconsistent/flexible, giving rise to complementary degrees of reactivity to occasions of overlap (Roberts & Goldstone, 2011). But why did one of the players become more stubborn? We found that if a player tends to select tiles consistent with a focal region (that is, one half of a focal split), they tend to be more consistent. In other words, the closer a player's tile selection strategy is to a focal region, the more stubborn they become, presumably because they believe that they are forming one half of a viable division of labor. The regression model of consistency with respect to distance to closest focal region confirms this effect ($\beta \approx -0.12$; $p < 0.001$). The interesting question now is how the other player figured out that they have to select tiles in the appropriate complementary region, given that a player only had access to their own uncovered tiles and not the other player's uncovered tiles. The key seems to lie in the fact that players do have access to overlapping tiles, from which the other player's selected tiles can be inferred with reasonably high validity.

One mechanism that can account for many players' shifts in selected tiles is based on a measure of the similarity between a focal region and the overlapping tiles. If one player's selected tiles are sufficiently close to a focal region, then this can be used as a signal for the other player to select the corresponding, complementary region. In Figure 3 we take a closer look at an actual game play from a dyad in which this mechanism is prominent, as exhibited by Player B's transition. On Round 23 the overlapping tiles are similar to the focal region RIGHT, which inclines B to select every single tile in the complementary LEFT region. Observe that B not only re-selected the left region's tiles from the previous round, but uncovered the whole LEFT region. More generally, the player's attention is attracted toward a focal region k when the region that is complementary to k is sufficiently similar to the overlapping tiles. To be sure, even though the process seems to be gradual and there are other factors at play, these complementary focal regions have attraction power. Last but not least, observe that the overlapping tiles are the same for both players, so Player A's attention is also attracted by LEFT. Nevertheless, given that A has uncovered the focal region RIGHT, they tend to become "stubborn" in the sense of resisting substantial change to their uncovered tiles. The combination of this retention and the attraction powers of a focal region informs a decision process that we call the Focal Regions as Attractors heuristic (FRA).

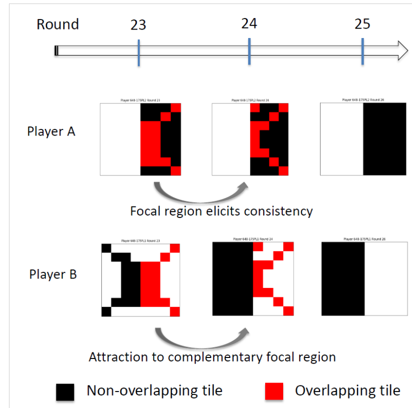

Figure 3: Evidence in favor of Focal Regions as Attractors (FRA). We see the transition from one round to the next, taken from an actual game play. In each grid, black tiles represent uncovered tiles and red tiles were uncovered by both players. Player A's transition illustrates 'stubbornness' and Player B's illustrates the attraction exerted by the complement of A's focal region, which is also a focal region. See details in the text.

Computational models

We put our previous explanations to the test by providing a computational model for each one of these two heuristics. The first model is an implementation of WSLS. To motivate it, suppose that on round n the player uncovered tiles determining BOTTOM. We want to determine the probability of choosing each region k in $\mathcal{K}$ on round $n+1$, where k in $\mathcal{K}$={RS, ALL, NOTHING, BOTTOM, TOP, LEFT, RIGHT, INSIDE, OUTSIDE}. $\mathcal{K}$ contains the focal regions, plus the type of region we call RS, which represents all remaining regions in the grid. Now, if the player is in a win situation, we should increase the probability of choosing again BOTTOM. This effect can be obtained by means of a threshold function (see Figure 4). More formally, the model defines a probability function, determined by the following formula:

$$P(k) = \frac{\text{attract}(k)}{\sum_{r \in \mathcal{K}} \text{attract}(r)} \tag{1}$$

The attract(k) function represents the extent to which a player is inclined to choose region k, given the current state of the game. For the WSLS model, we assume that this state is represented by the vector (i, s), where i is the region explored on the previous round and s the obtained score. The attract function for the WSLS model is defined in the following way:

$$\text{attract}(k, i, s) = \text{bias}_k + \alpha * \text{thresh}(s_n, \beta, \gamma) * \text{I}(k, i) \tag{2}$$

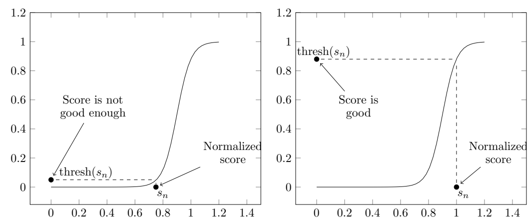

Figure 4: The thresh(s_n) function representing how good the score was. The left panel illustrates a situation where the score is not good enough, which is captured by the low value of thresh(s_n). The right panel illustrates a good situation, captured by the high value of thresh(s_n).

Here, the term bias_k represents how inclined the player feels toward k, all other things being equal, and is expected to be higher for more pre-experimentally salient regions. The second term contains the functions thresh and I, which are defined in the following way:

$$\text{thresh}(s_n, \beta, \gamma) = \frac{1}{1+e^{-\beta(s_n-\gamma)}}, \quad \text{I}(k,i) = \begin{cases} 1, & \text{if } i = k \neq \text{RS} \\ 0, & \text{otherwise} \end{cases}$$

Here, s_n is the normalized score, which takes values between 0 and 1. The function thresh(s_n, β, γ) has an S shape and takes values in the open interval $(0,1)$. It goes from values near 0 to values near 1 when s_n is near γ, and the steepness of this transition is determined by β (see Figure 4). The second term in Equation 2 contains the parameter α, which determines the extent to which the score increases the player's tendency to choose k, when the normalized score is greater than γ. The effect of I(k,i) in this expression is that the only region that has its bias modified is region i (i.e., the region explored on the previous round) and only if this region is a focal region. The value of attract(k) for the remaining regions is equal to bias_k.

The model defined by FRA extends the previous model. To motivate it, suppose that on round n the player uncovered tiles in the i region as defined in Figure 5. Now, we should consider the overlapping region, j, and consider its similarity to each focal regions (see right panel of Figure 5). The more similar to k, the more attractive *the complement* of k becomes. In our example, the overlapping region is more similar to UP, so the probability of choosing BOTTOM on round n+1 is increased. More formally, we assume that the current state of the game is represented by the vector (i,s,j), where i is the region explored on the previous round and s the obtained score, and j the area formed by the overlapping tiles. The attractiveness of k is defined in the following way:

$$\begin{aligned} \text{attract}(k,i,j,s) = \text{bias}_k + \alpha * \text{thresh}(s_n,\beta,\gamma) * \text{I}(k,i) \\ + \delta * \text{sim}(j,\bar{k},\varepsilon) * \text{Focal}(k) + \zeta * \text{I}(k,i) \end{aligned} \quad (3)$$

Observe that the first two terms in Equation 3 are the same as in Equation 2. The third and fourth terms are new. In the third term, the function sim$(j,\bar{k},\varepsilon)$ determines the similarity

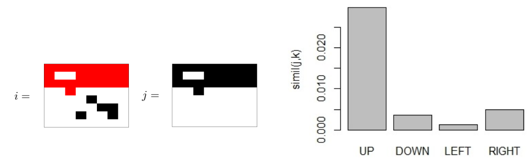

Figure 5: An example of a region visited, i, the overlapping tiles, j, and the similarity between j and other regions.

Model	θ_1	θ_2	θ_3	θ_4	α	β
WSLS	0.14	0.0674	0.0123	0.0009	39	405
FRA	0.077	0.048	≈ 0	≈ 0	48	402

Model	γ	δ	ε	ζ	Dev.	AIC
WSLS	0.933	0	0	0	3060	3074
FRA	0.99	1.57	0.94	3	2709	2709

Table 1: Best parameters and deviance for each model. The first four parameters correspond to the biases in the model: θ_1=bias_{ALL}, θ_2=$\text{bias}_{\text{NOTHING}}$, θ_3=$\text{bias}_{\text{BOTTOM}}$=$\text{bias}_{\text{TOP}}$=$\text{bias}_{\text{LEFT}}$=$\text{bias}_{\text{RIGHT}}$, and θ_4=bias_{IN}=bias_{OUT}. Moreover, bias_{RS} is defined as 1 minus the sum of the other biases, and we require that the sum of all biases adds to 1.

between j and the complement of k, denoted as $\bar{k}$. The function Focal(k) is defined in the following way:

$$\text{Focal}(k) = \begin{cases} 1, & \text{if } k \notin \{\text{RS, ALL}\} \\ 0, & \text{otherwise} \end{cases}$$

The parameter δ in Equation 3 determines the extent to which the similarity between j and $\bar{k}$ modifies attract(k), but this only occurs when k is a focal region and is different from ALL. This effect is obtained by multiplying δ by Focal(k). Finally, the parameter ζ determines the extent of the player's stubbornness when i is a focal region.

Note that the extra parameters from FRA with respect to WSLS are δ, ε, and ζ, and that Equation 2 for WSLS can be obtained from Equation 3 when $\delta = \zeta = 0$. That is, WSLS is a nested model within FRA.

Using maximization of log likelihood of the multinomial distribution of the observed transition frequencies and the respective predicted probabilities given by the model, we found the optimal parameters and the *deviance* of the two models, summarized in Table 1. Both the Likelihood Ratio Test ($\chi^2 = 351$; 3 d.o.f.; p< 0.001) and the ΔAIC=365 provide quantitative evidence that the additional parameters contributed by FRA provide a better account of the underlying choice process and that this model's better fit to the data is not due to overfitting.

We simulated our game in the same conditions as the experimental task. For each model, we ran 100 simulations of 60 rounds of the game, obtaining two collections of simulated data. In the two top panels of Figure 6 we can observe

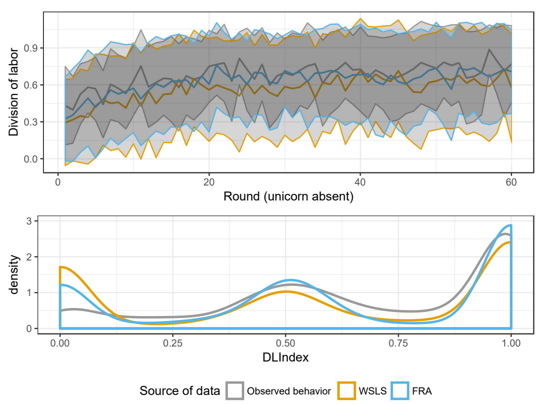

Figure 6: Comparison between observed and simulated behavior. The top panel shows the behavior of DLINDEX through all rounds. The vertical axis represents, for each round, the average DLINDEX. Shadow regions represent an error margin of one standard deviation. The bottom panel shows the kernel density estimate of DLINDEX for observed and simulated data, where each observation is the DLINDEX of a dyad at the end of a round.

the behavior of DLINDEX through all rounds. The vertical axis represents, for each round, the average DLINDEX with respect to all dyads in the respective set (45 for humans; 100 for each model). In the three cases we observe that the indexes are within the error margins of the others, and that the three sets of data show the same positive trend through rounds. However, in the case of WSLS, the t-test of mean difference (p≈ 0.01) does not provide conclusive evidence to assert that the means are the same, whereas the t-test of mean difference of DLINDEX (p≈ 0.3) determines that there is no statistical evidence to claim that the means are different, which means that FRA is better at capturing the tendency of DLINDEX in human subjects.

In the bottom panel of Fig. 6 we can see the kernel density estimate of DLINDEX for observed and simulated data. When the density curve is high (y-axis) for a given value of DLINDEX (x-axis), it means that there were many rounds for which a dyad obtained a DLINDEX close to x. Observe that, for humans, high values of DLINDEX are more frequent than medium and low values—representing the fact that many dyads split the grid satisfactorily. However, in WSLS there is a considerable tail on the left, indicating many more trials on which dyads did not split the grid into complementary regions, as compared to humans. Moreover, in WSLS the frequency of low values is higher than that of medium values, which is not in accordance with the observed data. For FRA, the frequency of low values is not greater than that of medium values, which is closer to what is observed in human data. To sum up, it seems that WSLS predicts a less efficient division of labor than exhibited by people, whereas FRA and people show a comparable degree of division of labor.

Discussion

57% of human dyads finished 60 rounds of game play with an efficient division of labor. The results from our experiment and our computational models allow us to explain how most dyads managed to split the grid without being able to engage in explicit negotiations. First of all, even though there are 2^{64} ways to split the grid, dyads split it in only four different ways. In some sense, these splits are focal points because they have a certain psychological salience (Schelling, 1960). One might have thought that these individual cognitive biases (focal points) would exert an early (in terms of rounds) influence on choices exactly because they are *a priori*, so that agents would have started on Round 1 with strategies of selecting all tiles on the left, top, bottom, or right. If agents understand that these are natural attractors, then through engaging in many levels of iterated thinking based on common knowledge (Lewis, 1969), these would be logical starting points. However, players do not generally start with strategies that resemble focal points. Humans are far more idiosyncratic and exploratory in their initial selections of tiles. It is only through repeated interactions that players manifest their *a priori* predispositions/biases toward certain focal points. In other words, a *priori biases* do not entail that the biases are manifest at the onset of play. It is only through dyadic interaction that these biases are revealed (Kaush ML, Griffiths TL, & Lewandowsky S, 2007). Returning to our two roommates example, there are 128 different ways to divide the days of the week in order to alternate one roommate cooking for two. We suppose that not every possible division is equally salient for them, and that only a handful of divisions will actually attract and retain their attention, such as the division between weekdays or weekend, or a division based on the idea of cooking every other day. If the roommates cannot explicitly negotiate a division but are given the daily chore of preparing lunch(es), one roommate will eventually follow one of these psychologically salient divisions and will tend to persist in the strategy because it is a focal point. To the extent that the other roommate wants to avoid overlapping days, soon they will be attracted to the psychologically salient strategy of choosing complementary days of the week. Interacting individuals, both human and algorithmic, can often arrive at efficient coordinating solutions in a paradigm that incorporates two challenging conditions – individuals cannot explicitly communicate, and there are multiple coordinating solutions that are initially equally salient. The human and computational results indicate that agents solve this coordination task by beginning with a set of possibly incompatible focal points. Then, via iterated interactions they adjust their behaviors to move toward focal points when they are not at a focal point, stay in a focal point once reached, and shift to a complementary focal point relative to the other player. In this way, the coordination that a group forms results from the interplay over time between their *a priori* cognitive biases and the dynamics of their interpersonal interaction (Hawkins, Goodman, & Goldstone, in press).

Acknowledgements

EAL was partially supported by the following research awards from Universidad del Rosario: "Becas para Estancias de Docencia e Investigación, 2018 and "Fondo de Capital Semilla, 2018". EAL is also grateful with Luis Andrade for useful discussions. Both authors are grateful with Evan Nix for running the experimental sessions, and with the anonymous referees at CogSci2019 for useful comments.

References

Arthur, W. B. (1994). Inductive Reasoning and Bounded Rationality. *The American Economic Review*, *84*(2), 406–411.

Balietti, S. (2017). nodeGame: Real-time, synchronous, online experiments in the browser. *Behavior Research Methods*, *49*(5), 1696–1715.

Colman, A. M. (2003). Cooperation, psychological game theory, and limitations of rationality in social interaction. *Behavioral and Brain Sciences*, *26*(2), 139–153.

Hawkins, R., Goodman, N., & Goldstone, R. (in press). The emergence of social norms and conventions. *Trends in Cognitve Science*.

Heylighen, F. (2013). Self-organization in Communicating Groups: The Emergence of Coordination, Shared References and Collective Intelligence. In *Complexity Perspectives on Language, Communication and Society* (pp. 117–149). Berlin, Heidelberg: Springer.

Holbrook, M. B. (2002). Bounded rationality: The adaptive toolbox. *Psychology & Marketing*, *20*(1), 87–92.

Kaush ML, Griffiths TL, & Lewandowsky S. (2007). Iterated learning: intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin & Review*, *14*(2), 288–288–94.

Kearns, M., Suri, S., & Montfort, N. (2006). An Experimental Study of the Coloring Problem on Human Subject Networks. *Science*, *313*(5788), 824–824–827.

Kennedy, J. F., Eberhart, R. C., & Shi, Y. (2001). *Swarm Intelligence*. San Francisco: Morgan Kaufmann.

Lewis, D. (1969). *Convention: A philosophical study*. Wiley-Blackwell.

Mehta, J., Starmer, C., & Sugden, R. (1994). The Nature of Salience: An Experimental Investigation of Pure Coordination Games. *The American Economic Review*, *84*(3), 658–673.

Roberts, M. E., & Goldstone, R. L. (2011). Adaptive Group Coordination and Role Differentiation. *PLOS ONE*, *6*(7), e22377.

Robinson, G. E. (1992). Regulation of Division of Labor in Insect Societies. *Annual Review of Entomology*, *37*(1), 637–665.

Ross, D. (2018). Game Theory. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018 ed.). Metaphysics Research Lab, Stanford University.

Schelling, T. (1960). *The Strategy of Conflict* (1st ed.). Cambridge: Harvard University Press.

Simon, H. A. (1957). *Models of man: social and rational; mathematical essays on rational human behavior in society setting.* New York: Wiley.

Sloman, S., & Fernbach, P. (2017). *The Knowledge Illusion: Why we never think alone*. Riverhead.

Smith, A. (2008). *An inquiry into the nature and causes of the wealth of nations Adam Smith*.

Weitekamp, C., Libbrecht, R., & Keller, L. (2017). Genetics and Evolution of Social Behavior in Insects. *Annual Review Of Genetics*, *51*, 219–219–239.