Kartik Tiwari

Prof. Sharon Elizabeth Berry

History of Analytic Philosophy

17 May 2022

Later Wittgenstein, Game Theory and Information Theory

1. Introduction

Wittgenstein in *Philosophical Investigations* makes a case for why attempts at formalization of language are inadequate descriptions of the true scope and variety of natural language[1]. Rather, the appropriate method of capturing the mechanics of human communication is - according to Wittgenstein - via the Theory of *Language Games*. In the context of language games, it becomes immediately clear that the meaning of a word varies with what language-game is being played at the time. Communication is a series of moves played by players who have mutually consented to a convention of rules. Wittgenstein himself gives several examples of such language games - some from use case scenarios of natural language and some toy models he invented - to discuss various features of communications. **In my paper, I would analyze Wittgenstein's Colored-Box Language Game (PI §48) and demonstrate how tools of game theory and information theory clarify some of his remarks.** In particular, with the Colored-Box language game, Wittgenstein was concerned with the following two ideas -

1. How does one learn the Colored-Box language? In the process of learning from others game-play, how does one effectively differentiate a correct move (syntactically

[1] To avoid strangling myself in lines of philosophical investigation which deviate from the central theme of the paper, I define *Natural Language* loosely as that system of symbols which allows the reader of the current text to comprehend its semantic load.

and semantically proper communication) from a mistaken one? (PI §53)

2. Are there rigorous grounds to differentiate simple and complex elements in the Colored-Box language? How many elements is a sentence made of? (PI §48)

In the process of addressing these, I would draw from the vast amounts of extant literature on evolutionary game theory applied to signalling games. However, prior to applying these tools from different domains to Wittgenstein's analysis, we must cast his language game in the formalism of Signalling games.

2. Casting Colored-Box Language Game Onto Signalling Games Formalism

Originally, Wittgenstein introduces the Colored-Box language (henceforth referred as CB) as descriptions of a complex of 9-squares which could each be colored red, green, black or white. But we can simplify this even further while preserving all of the philosophical content that was to be extracted from the use of this toy language. For the purposes of this paper, let us instead consider a 4-squares grid with the possibility of each tile being either white or black (see Fig. 1). The reason for this simplification is to make the problem more computationally tractable (although even in this form, we shall soon see how casting it into the signalling game formalism leads to an enormous amounts of calculations). For brevity, I shall refer to this modified colored box game as MCB.

2.1 Setting Up Signalling Game Formalism

Signalling games were envisioned by David Lewis in response to Willard Van Quine's skepticism about the origin of language. Quine's objection to the convention-view of language can be boiled down to the following observation - if language is a mutually agreed upon convention of symbols and sounds, then how was the agreement arrived at in absence of a pre-existing notion of semantics. Or, as Bertrand Russell remarks, 'We can

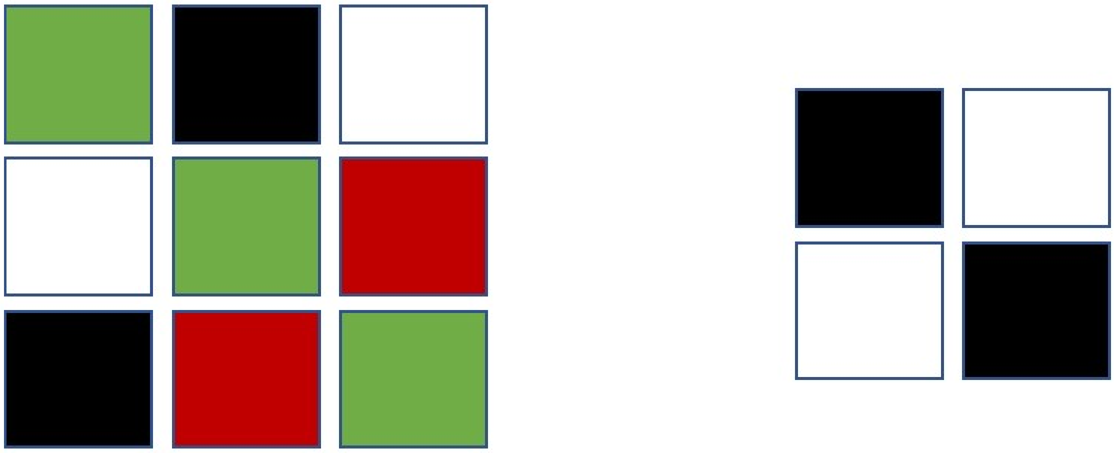

Figure 1: On the left, we have an example of a possible configuration in Wittgenstein's original colored-box game. On the right, we have a possible configuration of a simplified version of colored-box game.

hardly suppose a parliament of hitherto speechless elders meeting together and agreeing to call a cow a cow and a wolf a wolf'. Lewis, via the use of signalling games, demonstrates that 'an agreement sufficient to create a convention need not be a transaction involving language'.

I shall now proceed to set-up the most simple possible Signalling Game which involves two-players, two-states and two-signals. The principles demonstrated can then be easily generalized to our MCB language game. Consider the game is being played by a Sender (S) and a Receiver (R) of messages. Nature might randomly expose the Sender to two possible states - an incoming hawk t_1 or an incoming lion t_2 - based on which Sender might send a signal to the Receiver. The Receiver, interprets the message received and chooses one of two possible actions - hide in a cave a_1 or climb a tree a_2. The vocabulary of S consists of two possible messages (m_1 and m_2). The nature of these messages is irrelevant (for example, the messages can be some strings with phonetic structures or gestures resembling the characteristics of a lion or hawk) as long as the two messages can be differentiated from each other. The signalling game can be summarized as follows -

S maps the set of observable states of nature to the set of available messages (let us call this Sender's map $\rho : T \to M$) and transmits it to R, who then maps the set of possible messages to the set of possible actions (and let us call this Receiver's map $\sigma : M \to A$). A combination of possible sender map and reciever map together form a communication strategy (ρ, σ).

Clearly, for each state of nature there is a corresponding correct choice of action i.e. to protect oneself from a lion one should climb a tree and to protect oneself from a hawk one should hide in a cave. Therefore, in the terminology of game theory, we can create a utility table (Fig. 2) which demonstrates that the choice of actions incompatible with nature's states are negatively incentivized for the players. At this juncture, it is important to remind ourselves that signalling games are methods of modelling features of language. We must not concern ourselves with individual numbers moving forward but only with trends and patterns of the mathematical model which are invariant to our arbitrary choice of numbers. Therefore, what matters in Fig. 2 is not what numbers are kept at each position in the table but the general observation that some choices are inferior to others.

	a_1	a_2
t_1	0	-1
t_2	-1	0

Figure 2: Negative pay-off for action choices incompatible with nature's states

If we enumerate all possible ρ maps and σ maps (i.e. all ways of transcribing the information of nature's states into signals and all ways of interpreting signals into actions) then we can enumerate all communication strategies. For our simple 2-state 2-signal 2-action game, we get 16 possible communication strategies (4 possible ρ maps and 4 possible σ maps which can be combined with each other in 16 possible pairs). For each such communication strategy (ρ_i, σ_j), we can consider the expected pay-off for all states and this can help us gauge how these communication strategies perform with respect to each other (Fig. 3). The mathematical details [2] of similar (but not same) models can be found in Skyrms 2010 and Correria 2019.

	σ_1: $m_1 \to a_1$, $t_2 \to a_1$	σ_2: $m_1 \to a_1$, $t_2 \to a_2$	σ_3: $m_1 \to a_2$, $t_2 \to a_1$	σ_4: $m_1 \to a_2$, $t_2 \to a_2$
ρ_1: $t_1 \to m_1$, $t_2 \to m_1$	s_1 | -1	s_2 | -1	s_3 | -1	s_4 | -1
ρ_2: $t_1 \to m_1$, $t_2 \to m_2$	s_5 | -1	s_6 | 0	s_7 | -2	s_8 | -1
ρ_3: $t_1 \to m_2$, $t_2 \to m_1$	s_9 | -1	s_{10} | -2	s_{11} | 0	s_{12} | -1
ρ_4: $t_1 \to m_2$, $t_2 \to m_2$	s_{13} | -1	s_{14} | -1	s_{15} | -1	s_{16} | -1

Figure 3: Expected pay-offs for each of the 16 possible communication strategies

Notice how strategies s_{10} and s_7 are particularly horrible ways of playing this signalling game as for each message, the reciever does the opposite of what the appropriate action would be. On the other hand, s_6 and s_{11} would be the best strategies for playing the game because then the reciever does precisely the correct corresponding action for each message received.

2.2 Generalizing to MCB Language Game

We can perform a similar treatment for our MCB language game as well to cast it in the signalling games formalism. Except, now, the total number of states[3] (i.e. configurations of colored tiles in the grid) is 16 and we have 16 corresponding messages available to us (most trivially in the form of strings such as WWWW, WBWB, WBBB, etc.). Considering there are still two players playing this language game, we will have 16^{16} possible ρ maps and 16^{16} possible σ maps. This, in effect means, there are $(16^{16})^2 = 16^{32}$ strategies to play this game. Though calculating the expected pay-off for each strategy in

[2] I wrote a small python code to populate the table in Fig. 3

[3] 2 possible states for 4 possible locations, which is the same as saying $2^4 = 16$.

this game would be computationally intractable due to the sheer size, there is in principle nothing conceptually different in MCB compared to the simple two-state two-action game described earlier. Therefore, while I shall not perform explicit calculations for MCB, we can rest assured that the qualitative features of numerical investigations of the simple signalling game would carry over to MCB too.

Now that the foundations of the signalling game formalism have been established, we can revisit the two aforementioned remarks by Wittgenstein on learning and on identifying simple units of CB language game.

3. Evolutionary Dynamics and Learning

As established in the previous section, certain communication strategies have inherently more value than other ones depending on the context in which the game is being played. But expecting the players to know the pay-off table at the beginning of the game in order for each player to settle on which strategy to employ is an extremely big ask. Recall, even in the fairly simple case of MCB, the total number of strategies was of the order of 16^{32}. Fortunately, for emergence of meaning, in Lewis's own words 'we only require a propensity to conform to a regularity'. In the process of unpacking this statement, we will also arrive at a plausible response to Wittgenstein's inquiry about an organic method by which one could learn the CB language game.

Wittgenstein appeals to various ways in which the rules of CB game can be acquired by a player (PI §54). I categorize his rule-acquisition techniques under two categories - explicit transfers and implicit transfers. Explicit transfers involve the learner being told what the rules are or being asked to practice the game under supervision and implicit rule transfers take place when a learner tries to abstract away the rules of the language game by simply observing two other players playing the game. The problem with explicit rule transfers is that, as mentioned earlier, in case of absolutely no shared pre-existing notion of semantics there is no effective vessel for facilitating rule transfers[4]. On the other hand, the problem with implicit rule transfers (as Wittgenstein himself points out) is how would a learner distinguish a valid move from a mistaken one? For Wittgenstein, this is can be done by looking for other non-verbal cues which humans employ when they commit a mistake (such as a particular head nod after a slip of tongue). However, considering that the non-verbal cues are also a form of language-game, we will now have to explain how one learns this different language-game - which just begs the question.

At this point, we interject to introduce Darwinian evolutionary dynamics (also called replicator dynamics by some authors) in our game-theoretic formalism. The essential premise is that players play many iterations of the signalling game. If we refer back to our simple Hawk-Lion signalling game, we have an extreme example of how players who choose effective communication strategies have a higher chance of survival in subsequent iterations of the game. If we assume that initially in a population group all communication strategies are equiprobable (or have any random probability distribution) then by observing how the population fractions evolve over time, we will notice that the those communications strategies which are effective would quickly spread through the population. We can model this by setting up a simple differential equation which connects the rate of change of a fraction of population x_i which follows the communication strategy s_i with a function η of that strategy's expected pay-off.

$$\frac{dx_i}{dt} = \eta(s_i)x_i \tag{1}$$

[4]For example, one might try to point to certain configurations and then to certain strings to demonstrate some kind of correspondence between the two. However, this presumes that the learner will necessarily grant some semantic significance to the physical gesture of 'pointing'.

In the simple case of Hawk-Lion Signalling game, we have three different possible pay-offs: 0, -1, -2 (see Fig. 3). In our natural selection interpretation of the replicator dynamics, we shall call these pay-offs 'Strategy Fitness'. On solving Eq. 1 (on python) for a population in which it is equiprobable for any player to adopt one of the three strategies, we get populations dynamics displayed in Fig. 4.

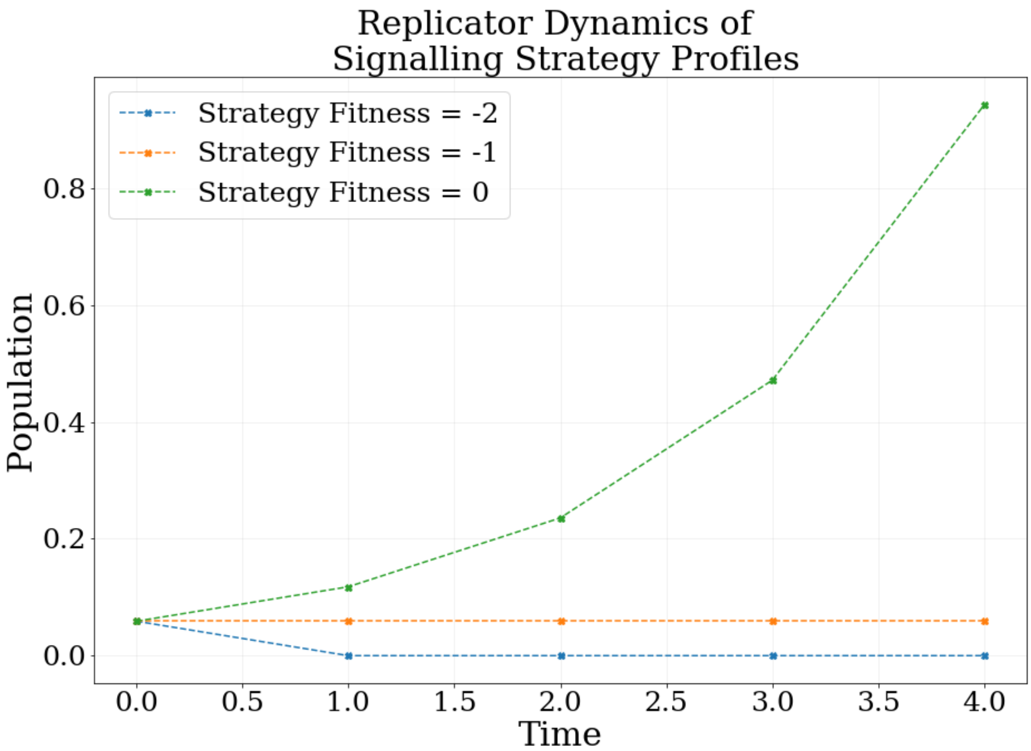

Figure 4: Numerical results demonstrating that the most effective communication strategy, over multiple iterations of the signalling game, becomes more prevalent than the ineffective strategies.

We need not necessarily consider such a natural selection based fitness interpretation for our replicator dynamics. We can also construct a model where certain strategies are positively incentivised. Further, there is a vast amount of existing literature on various learning strategies (such as Roth-Erev Reinforcement, Bush-Mosteller Reinforcement or other more complex reinforcements) that vary different parameters of our system to achieve various kinds of learning curves. However, the answer to the question of whether learning of meaningful signalling strategies without ostensive instructions is possible under suitable situations is affirmative. To re-emphasize, this emergence of meaningful signalling is guaranteed due to an asymmetry in the inherent value of communication strategies granted to them by courtesy of their varying compatibility with their environment. Since, in principle, all that differs between the Hawk-Lion signalling game and MCB language game is the number of states, messages and actions, we can conclude that, under suitable environments, learning of the MCB game (and subsequently CB game) can happen even without the existence of any pre-existing shared semantic grounds.

In this framework, what Wittgenstein calls a mistake just forms one of the other possible communication strategies for CB language game. The identification of a mistake, under these replicator dynamics, need not be the responsibility of the learner themselves but can be positively or negatively incentivized by the external environment. For example, one can envision two players playing the CB language game and every time the information about the state of the complex is correctly interpreted by the reciever, an external third-party pays both the players Rs. 100. Now, instead of having to rely on another non-verbal language game, a learner employing an implicit transfer of rules can trivially identify a mistake by simply keeping track of the pay-offs of the players at the end of each round of communication. Note again that this is made possible because the environment itself prefers certain communication strategies over others.

A possible concern here, for me, is that considering how computationally intractable the model becomes even for the simple MCB game, it seems unlikely that such learning mechanisms are followed to acquire the exponentially more sophisticated natural language. I could not find a response to this objection in the existing literature on signalling games. My (admittedly loose) response to this observation would be an appeal to the fact that the entirety of natural language need not have emerged via such evolutionary mechanisms. The need for replicator dynamics is to explain the emergence of convention in absence of pre-existing semantics. However, once enough tools of language have been assembled and developed in order for people to talk *about* language, ostensive instructions and modifications of language become possible - accelerating the acquisition of linguistic prowess for a community manifold.

4. Dismissal of Absolute Simples

Now, let us move our focus to Wittgenstein's remarks on the nature of simple and complex elements in the CB language game. Wittgenstein contests the idea that in the CB language game, the simplest element of the language would be a single letter of a string which corresponds to a particular arrangement of colored tiles (such as 'W' in 'WWRGGRWRW'). He instead argues that what counts as 'simple' also depends on what language game is being played. Let me clarify this more explicity with respect to what this would mean in our modified color box game. In MCB, we have 16 possible states which can be presented to the sender of the signal (as depicted in Fig. 5).

One way to carve out these natural states is, obviously, in terms of 4 slots which can each be filled by two possible characters; the total number of possible states becomes $2 \times 2 \times 2 \times 2 = 2^4 = 16$. However, one can equivalently carve out these states of nature in terms of two vertical (or horizontal) rectangles which can each be represented by 4 possible characters 'A', 'B', 'C' and 'D' (i.e. a total number of $4 \times 4 = 4^2 = 16$ possible states). The two ways of playing the MCB language game when presented with the same set of possible states is displayed in Fig. 6. Notice how the strings 'BBWB' and 'CD' represent the same states but with different grammars. In the second way of playing the

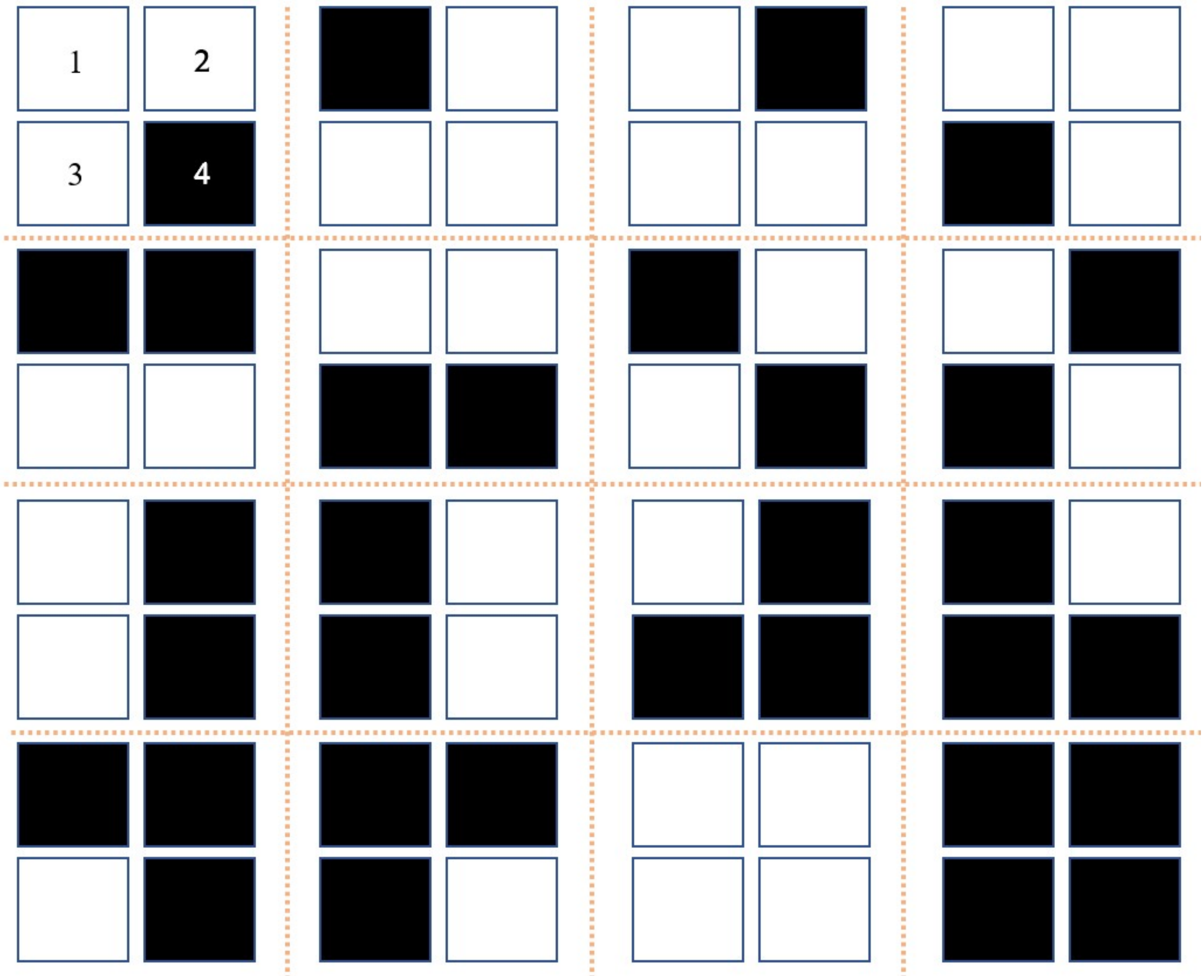

Figure 5: Possible states that can be presented by nature to the sender of signals

MCB game, one would then naturally concede that the simplest unit of the complex is a vertical section of the grid.

Which representation of MCB is more fundamental is a question which lacks a clearly defined answer from Wittgenstein's point of view. There is one framework, however, in which the 'elementariness' of a unit of language can be gauged unambiguously i.e. Claude Shannon's *Information Theory.* Whether or not this metric has any philosophically useful metaphysical and epistemological application in studying language can be questioned[5]. Regardless, ideas from information theory do play a significant role in the model-building

[5]While some philosophers strongly prescribe that epistemology must reorient itself to become a study of flow of information, others believe that the *intentionality* of 'real' information is utterly incompatible with the mathematical theory of information.

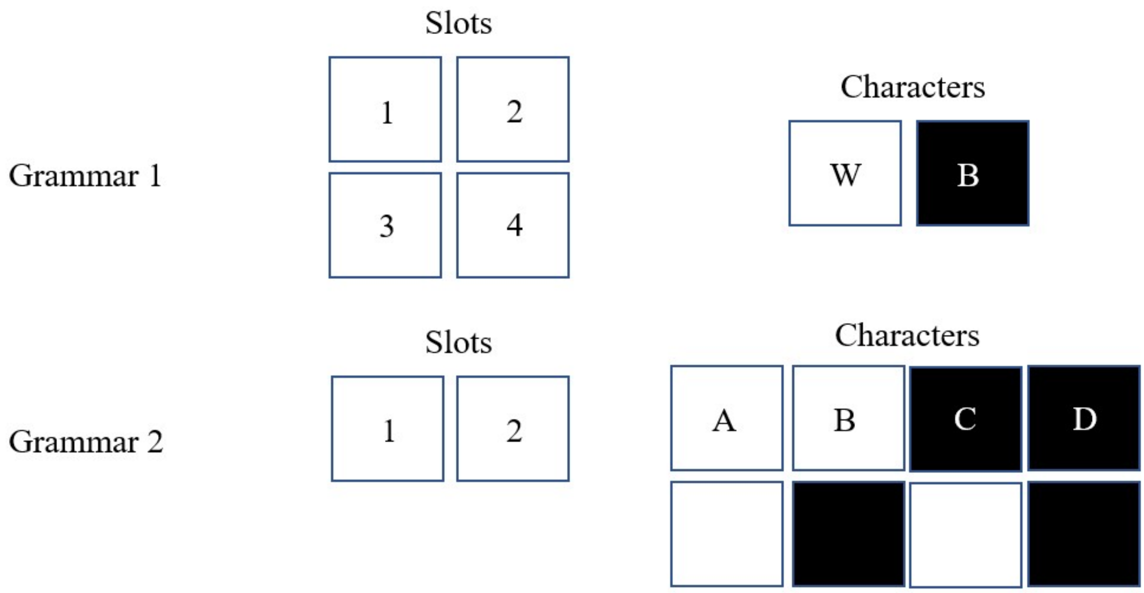

Figure 6: Two possible ways of playing MCB language game as either a 4-slot 2-character game or 2-slot 4-character game.

process of signalling games. Thus, irrespective of its perceived philosophical utility, I shall now proceed to demonstrate how concepts of information theory can help disambiguate simple units from complex ones.

In Information Theory, the information content of a message is thought of in terms of probabilities. As an edge case, consider a scenario where the sender and reciever both know an event is certain and inevitable. In such a scenario, a message containing a declaration of that event's occurrence (as per Information Theory) carries no information at all. Because details and discussions of Information theory are ubiquitous and easily accessible, I shall not try and reconstruct the foundations of information theory in this paper. Instead of appealing to the formula of self-information of a state[6] in information theory, I would argue on intuitive principles why one of the two ways of playing MCB language game displayed in Fig. 6 can be thought of as utilizing simple elements. To do so, I first define the 'simplest' element as that element of the language-game which

[6]which is $I(x) = -\log_2 P(x)$ where P(x) is the probability of system being found in state x

carries the smallest amount of information.

Recall that in either of the ways of playing MCB, there are 16 possible states. If we assume each state to be equiprobable, then any single state has the probability 1/16 of occurrence. The information content of a complete message would rise this degree of certainty from 1/16 to 1 - while an incomplete message would rise the certainty only partially. In Grammar 2, if we know the first letter of the string that corresponds to a state, the degree of certainty about the system being in a certain state rises from 1/16 to 1/4 (because there are only 4 possible options for the second slot). On the other hand, in Grammar 1, if we know the first letter of the string then the degree of certainty rises from 1/16 to 1/8 - which is a smaller increase. This means that one character of Grammar 1 carries lesser information about the state than one character of Grammar 2. This is to be expected since each character of Grammar 2 is composed of two slots from the perspective of Grammar 1[7]. Therefore, a character of Grammar 1 is more simple than a character of Grammar 2 because it carries less information content.

One might interject here and remark that my metric for 'simplicity' of a unit is incommensurable with Wittgenstein's notion of 'simplicity'. I agree. Therefore, to substantiate Wittgenstein's dismissal of absolute 'simple units' one must first differentiate the different grounds on which one could possible talk about the idea of 'simplicity'. One possible way of talking about simplicity is whether in a particular language-game, a description of a state can be broken into syntactically smaller bits. Under such a definition, Wittgenstein's observation about the absence of a class of 'simple elements' invariant to carving out states differently holds true. However, if simplicity is to be defined by the information content of syntactically smallest units, then there is an effective way of comparing

[7]In technical terms, we would say that a character of Grammar 1 carries 1 bit of information while a character of Grammar 2 carries 2 bits of information

simplicity across two language games.

5. Conclusion

In this paper, I discussed how modern tools of Game theory and Information theory can be used to study Wittgenstein's paradigm of Language Games. In reference to Wittgenstein's colored-box language game, I demonstrated how learning is possible without ostensive instruction. I also demonstrated, for a precisely defined Information-theoretic sense of 'simplicity', Wittgenstein's dismissal of absolute simples breaks down. Remarkably, the traffic between Signalling Games and Wittgenstein's philosophy runs both ways. Not only can the theory of signalling games guide our efforts in understanding Wittgenstein but also Wittgenstein's methodological prescriptions and meta-philosophy serve as helpful cautions to avoid over reliance on artificial formalizations while model-building.

Works Cited

1. Barrett, Jeffrey A., and Jacob VanDrunen. "Language Games and the Emergence of Discourse." Synthese, vol. 200, no. 2, 2022, https://doi.org/10.1007/s11229-022-03645-7.

2. Benz, Anton. Language, Games, and Evolution Trends in Current Research on Language and Game Theory. Springer, 2011.

3. Darwin, Charles. The Origin of Species. D. Appleton, 1900.

4. Lewis, David. Convention a Philosophical Study. Harvard Univ. Press, 1974.

5. Correia, Jose. "Meaningful Play - Signaling Games in Light of Later Wittgenstein", 2019.

6. Parthasarathy, K. R. Information Theory. Research and Training School, Indian Statistical Institute, 1961.

7. Skyrms, Brian. Signals Evolution, Learning, Information. Oxford University Press, 2013.

8. Wittgenstein, Ludwig. "Philosophical Investigations". Wiley-Blackwell, 2010.