Old and New Riddles on Concept Sharing

Emanuele Bottazzi, Claudio Masolo, and Daniele Porello

Laboratory for Applied Ontology, ISTC-CNR, Trento, Italy

Abstract. We ask whether social interaction demands sharing social concepts. We illustrate our point by depicting possible situations that emerge when two individuals play chess. We formalize our hypothesis in First Order Logic and we show that the very idea of sharing social concepts poses an interesting challenge both from the standpoint of knowledge representation and of philosophical conceptual analysis. By endorsing a minimal notion of interaction, we conclude that sharing social concepts is not necessary for social interaction. Then, we relate our view to Wittgenstein's and Kripke's "Rule-following Considerations".

1 Introduction

Many authors views social concepts as depending on human beliefs and intentions [10,12] and, on this basis, some have argued that we have a certain form of "epistemic privilege" with regard to them, protecting us from certain possibilities of ignorance and error. We could be mistaken about the real nature of gold, as we were for centuries, but we could not be mistaken about the nature of marriages, parties or chess pieces, since they are human products, defined by us via collectively accepted rules [11]. Social concepts are strictly tied to rules [3], that is, rules "create the very possibility" ([10]: 27) for social, or more specifically, institutional activities—such as marriages, parties or chess games—to take place. This view promotes a tight connection between social interaction, concept sharing and rule following. For instance, in order to play chess, we need to grasp the concepts involved in a chess game (moves, pawns, capture, etc), that is, we need to accept the rules that define chess concepts. In this paper, we want to take a closer look at such an entanglement. In particular, we want to understand what does it mean to share social concepts and to what extent sharing is the grounding of interaction.

We illustrate our questions by means of the following scenarios. Two persons start playing chess. Black moves after White and so on. They both know the rules of chess and this is reflected by the fact that each agent acknowledges the move of the other as a correct move. That means that the moves are done according to the rules of chess. However, such an example represents only a very specific type of social interaction. Consider now two persons that plays chess possibly for the first time. They are not sure about the rules, for instance they may have a partial grasping of the rules that define the pawns, they may make mistakes, they may change their minds during the game. For example, White moves the rook as a queen, Black acknowledges this move because she is not certain of what the rules are, and they may keep on playing by adjusting their beliefs concerning the rules. We may question whether the two persons are still playing chess, however it seems hard to question that they are interacting. In this case, players may share very little of the rules of chess, hence they may share very little of the concepts involved in the chess game. We believe that this situation covers a broader spectrum of possible social interactions then any by the book chess game.

The question we want to approach is: What is it that two persons share whenever they interact? We will assume that social concepts are at least partially defined by rules, that is, the concept of a rook is at least partially defined by the admissible moves that a rook can make. In order to simplify our analysis, we will not discuss issues related to communication between interactants. Moreover, we concede that individuals have the same access to the physical or non-social reality, in order to exclude forms of disagreement that are not relevant to our point. For instance, players can recognize unambiguously the positions and the shapes of the different pieces of wood, they know the size of the board, and they understand that those pieces of wood stand for chess-pieces. We will discuss what does it mean to share social concepts by representing the players' views of the chess game in First Order Logic and by formalizing a number of hypothesis on social concepts. We will see that unpacking the very notion of sharing social concepts is quite involving and that a number of delicate ontological choices have to be made. We shall conclude that in many chess-like interactions, it appears that the two players may endorse very different chess concepts. That is, sharing social concepts appears not to be necessary for interaction.

We are assuming a minimal definition of interaction that is at least capable of accounting for cases such as our weird game of chess. This minimal form of interaction is intuitively a form of commitment to respond to the other's actions. For instance, White is interacting with Black because, whenever White moves, Black is committed to respond. Even in case White's move violates the rules that Black believes are holding, Black can in principle still respond by revising her rules. This notion of interaction is related to the notions of commitment and entitlement proposed by Robert Brandom [2]. The remainder of this paper is organized as follows. In Section 2, we present our formal treatment and a number of examples of chess interactions. Section 3 draws some connections between our points and Kripke's interpretation of Wittgenstein's 'Rule-following Considerations' in his *Philosophical Investigations*.

2 The illustrative example of chess formally illustrated

We start from the descriptions given by the two players of what happened during the game, i.e., from two subjective *reports* of the course of the game. We represent these reports by means of two FOL-theories—Π_{W} for White and Π_{B} for Black—describing how (and according to which rules) the chess moved and the way these moves are grounded on the *physical*-level, i.e., the level of bits of wood, stones, etc. as opposed to the *game*-level, i.e., the levels of rooks, bishops, etc. The physical-level is assumed to be *shared* by the players, i.e., while the players can have different viewpoints on the game, they agree on what happened in the physical world. This is a simplification hypothesis, one could think that also the physical-level is prone to disagreement.

The shared physical-level includes times (TM), physical objects (OB), e.g., bits of wood, stones, etc., positions on the board (PS), an *existence* predicate defined on both physical objects and positions—$\mathsf{E}(x,t)$ stands for "x exists at time t", and a *location* predicate—$\mathsf{L}(x,p,t)$ stands for "at time t, the physical object x is located in position

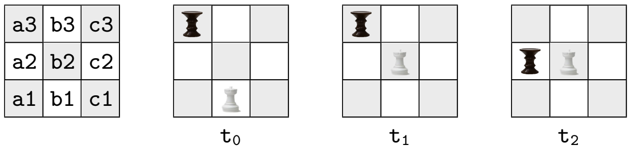

Fig. 1. Position labeling and three snapshots of the physical world.

p", where L, as all the predicates with a temporal argument, satisfies (a1) (where R is a generic relation that needs to be substituted by L in this case) and is functional, i.e., at a given time, the position of an object is unique (a2). The physicality of objects is further characterized by (a3), i.e., existent objects are always located. Furthermore, the 'board' exists during the whole game without changing, i.e., the positions are eternal (a4). For our aims an extremely simple model of the time of the game, the time of the moves, is enough: time is discrete, finite, linear, and left-bounded. We indicate by $t{+}1$ the (immediate) successor of time t and by t_{i+1} the (immediate) successor of t_i.

a1 $R(x_1,\ldots,x_n,t) \rightarrow \mathsf{E}(x_1,t) \wedge \cdots \wedge \mathsf{E}(x_n,t)$

a2 $\mathsf{L}(x,p,t) \wedge \mathsf{L}(x,p',t) \rightarrow p = p'$

a3 $\textsc{ob}(x) \wedge \mathsf{E}(x,t) \rightarrow \exists p(\mathsf{L}(x,p,t))$

a4 $\textsc{ps}(x) \rightarrow \mathsf{E}(x,t)$

The set of sentences that describe the existence and the position of the physical objects during the game—the report about what happened in the physical world—is included in both Π_{W} and Π_{B}. Focusing on a 3×3 fragment of a chessboard and labeling the positions as in Figure 1, the three snapshots of the physical world in Figure 1 are represented by (f1), where ♜ and ♖ are constants for physical-objects: OB(♜) and OB(♖).

f1 L(♜,a3,$\mathtt{t}_0$) $\wedge$ L(♖,b1,$\mathtt{t}_0$) $\wedge$ L(♜,a3,$\mathtt{t}_1$) $\wedge$ L(♖,b2,$\mathtt{t}_1$) $\wedge$ L(♜,a2,$\mathtt{t}_2$) $\wedge$ L(♖,b2,$\mathtt{t}_2$)

Another shared aspect regards who is moving. We do not explicitly represent neither the moves nor the player that is responsible for the transition between two configurations of the world. We just assume that White starts the game, she is responsible for the transaction from t_0 to t_1, and that the players correctly alternate. For instance, in the example in Figure 1, starting from an initial configuration at t_0, White does the first move, she moves ♖ in b2, while Black does the second move, she moves ♜ in a2.

The game-level includes chess pieces (CH), an extension of the existence predicate to chess-pieces, and a *manifestation* relation—$\mathsf{M}(x,c,t)$ stands for "at time t, the chess piece c is *manifested*, is *realized*, by the physical object x"—that satisfies (a1), (a5), i.e., M is *complete* manifestation, (a6), i.e., chess-pieces always have a manifestation, and, therefore, a location (through the location of its manifestation, see (d1)), and (a7), i.e., our model focuses only on the physical world involved in the game, objects are always manifestations of chess-pieces (not necessarily of the *same* piece). The object x is the physical substratum of the chess-piece c that has a more 'abstract' level of existence: c can survive a change of its substratum, i.e., its re-identification through time is not based solely on its physical manifestation. The idea is that the destruction or the damage of the manifestation of a chess-piece does not necessarily compromise the continuation of game, a replacement could be accepted by the players (see Example 2).

a5 $\mathsf{M}(x,c,t) \wedge \mathsf{M}(x',c,t) \rightarrow x = x'$

a6 $\textsc{ch}(c) \wedge \mathsf{E}(c,t) \rightarrow \exists x(\mathsf{M}(x,c,t))$

a7 $\textsc{ob}(x) \wedge \mathsf{E}(x,t) \rightarrow \exists c(\mathsf{M}(x,c,t))$

d1 $\mathsf{L}(c,p,t) \triangleq \exists x(\mathsf{L}(x,p,t) \wedge \mathsf{M}(x,c,t))$

On the one hand, the manifestation relation could be seen as a sort of existential dependence or, more specifically, a *material constitution*. On the other hand, one could think to a sort of *denotation* or *reference*, i.e., one could understand a chess-piece as a name, a definite description, an *individual concept* (a 'variable individual concept' [1]), or, more formally, as a (partial) function from times (possible worlds) to physical objects (individuals). In both cases, chess-pieces, e.g. "the white queen", are distinct from chess-types, e.g., "queen": the white queen is a queen. We represent the pieces as individuals (not as predicates subsumed by chess-types) for two reasons: (i) to be neutral with respect to the two different views on pieces and (ii) to simplify the model: by representing pieces as predicates, to understand whether or what piece moved one needs to explicitly provide a re-identification criterion that can be leaved implicit in our proposal.[1] Chess-types are represented by means of predicates like Queen: e.g., $\mathsf{Queen}(\mathtt{wq}) \wedge \mathsf{Queen}(\mathtt{bq})$, where wq and bq are individual constants naming chess-pieces, stands for "the white queen wq and the black queen bq are both queens".

As said, the theories of players Π_{W} and Π_{B} agree on the physical-level, i.e., they share both the *terminological* and *factual* (*assertional*) knowledge about the physical world, i.e., physical objects exist and are located objectively. The subjective perceptual capabilities of the players do not play any role at this level. Furthermore, Π_{W} and Π_{B} both adopt the previous axioms about M, the players agree on the meaning of M. However their factual knowledge about what (types of) pieces exist and what are their manifestations can differ, i.e., the game-level is subjective. For instance, players could disagree on the 'interpretation' of ♜ in the first snapshot in Figure 1, i.e., $\mathsf{M}(♜, \mathtt{bq}, \mathtt{t_0}) \in \Pi_{\mathrm{W}}$ whereas $\mathsf{M}(♜, \mathtt{bk}, \mathtt{t_0}) \in \Pi_{\mathrm{B}}$, where bq stands for 'the black queen' and bk for 'the black king'. In the following we provide some arguments for the view that White and Black can play chess without 'sharing' any chess-entity. For this reason we prefer to assume individual constants for chess-pieces to be local for each report, e.g., $\mathtt{bk_W}$ represents "the black king" according to White while $\mathtt{bk_B}$ represents "the black king" according to Black. Note that this is a formal move, it does not imply an ontological commitment on the privateness of chess-pieces, the individual constants must be seen here just as 'names'.[2]

The previous framework allows to represent the time-sequences of configurations of the objective physical-world and of the subjective game-worlds of the two players. Still,

[1] Conceiving pieces as individual concepts is a classical reification move, see for instance [9].

[2] The ontological dichotomy between a private, mental, or psychological nature vs. an abstract or social one concerns the theories of concepts in general. Note that privateness "doesn't preclude the sharing of a mental representation, since two people can have the same type of mental representation (...) When someone says that two people have the same concept, there is no need to suppose that she is saying that they both possess the same token concept." ([7], p.7)

the *rules* used by the players to accept these configurations as 'valid' (that allow the players to continue the game) are not in the model. Rules are often seen as what enable us to interpret what is going on in the brute (physical) reality as a game (institution).[3] Different kinds of rules can be distinguished, we will limit ourselves to rules that define the way chess-pieces move. In particular, we represent rules as *check constraints* that apply to chess-*types*, i.e., all the pieces of the same type are submitted to the same contratins, they move in the same way.[4] Chess-types can be used to formalize rules by means of sets of constraints on successive configurations of the game-world, i.e., for each chess-type T there is a set of necessary conditions with form (f2). The meaning of (f2) is: if at t a piece c of type T is located in the position $\mathtt{p_1}$, in the presence of one piece c' in location $\mathtt{p_2}$, then, at $t{+}1$, c cannot be located in (cannot move to) the position $\mathtt{p_3}$ ($\mathtt{p_1}$, $\mathtt{p_2}$, and $\mathtt{p_3}$ are individual constants).[5]

f2 $T(c) \rightarrow (\mathsf{L}(c, \mathtt{p_1}, t) \wedge \exists c'(\mathsf{L}(c', \mathtt{p_2}, t)) \rightarrow \neg\mathsf{L}(c, \mathtt{p_3}, t{+}1))$

We assume a sort of 'rationality' of players. Players stop to play if they are not able to 'solve a rule contradiction' manifested in a physical move. This does not mean that players cannot be 'flexible'. This simply means that when a player detects a contradiction, either she stops to play or she 'accepts' this contradiction by revising her rules or her factual game-knowledge to accommodate the contradicting move. This is motivated by our notion of interaction that is limited to the acceptance of moves, the others' responses must be taken into account to continue to interact. Flexibility could then mean (i) a revision of rules, i.e., in our framework, the substitution of a set of (f2)-constraints for T with a non logically equivalent new set for a different type T';[6] or (ii) a revision of the factual game-knowledge, e.g., the 'death' of some pieces and the 'birth' of new pieces, or a change in the manifestation of a given piece. One has then the problem of understanding if a chess-piece can survive a revision of rules, i.e., if migrations from T to its revision T' are allowed. Because chess-types are non temporally qualified, chess-types are implicitly essential for pieces, i.e., the identity (and re-identification) of pieces is based on how they move. To analyze type-migration, a temporal qualification of types is then necessary: $T(x, t) \wedge T'(y, t)$ could or could not be constrained to imply $x{=}y$. The possibility of type-migration requires to modify (f2) as in (f3).

f3 $T(c, t) \wedge T(c, t{+}1) \rightarrow (\mathsf{L}(c, \mathtt{p_1}, t) \wedge \exists c'(\mathsf{L}(c', \mathtt{p_2}, t)) \rightarrow \neg\mathsf{L}(c, \mathtt{p_3}, t{+}1))$

The fact that chess-pieces can survive a migration from T to T' highlights a *conceptual change* of the player, e.g., the player changes her mind regarding what queens are, she revises the rules for queens, i.e., intuitively, both T and T' individuate rules for queens. A conceptual change that shifts T to T' from t to $t{+}1$ impacts all the instances

[3] On that the literature is vast, starting from [13]. Often such rules are called *constitutive* [10].

[4] This does not hold for captures. For instance, a black rook in `a1` can capture a white queen in `b1` (and then move to `b1`). Vice versa a white rook in `a1` cannot move to a position occupied by a white queen. Because of the illustrative role of our model we do not consider captures.

[5] The condition $\exists c'(\mathsf{L}(c', \mathtt{p_2}, t))$ could be 'empty' (it is enough to instantiate it as a redundancy $\exists c'(\mathsf{L}(c', \mathtt{p_1}, t)))$, i.e., there could be positions that the Ts cannot reach (in one move) independently of the positions of the other pieces.

[6] Alternatively one could relativize (f2) to times as (where $\mathtt{t_1}, \ldots, \mathtt{t_m}$ are individual constants)

$$\mathsf{T}(c) \wedge (t = \mathtt{t_1} \vee ... \vee t = \mathtt{t_m}) \rightarrow (\mathsf{L}(c, \mathtt{p_1}, t) \wedge \exists c'(\mathsf{L}(c', \mathtt{p_2}, t)) \rightarrow \neg\mathsf{L}(c, \mathtt{p}, t{+}1))$$

of the concepts involved, all the instance, at t, of T migrate, at t', to T'. Vice versa, chess-pieces cannot survive a *misinterpretation*—e.g., the player confuses a T'-piece with a T-piece. To account for the distinction between conceptual change and misinterpretation, we group chess-types concepts into syntactically distinguishable classes of concepts. This is done by grouping the predicates of our language. Let $\mathcal{C}$ be a set of indexes and $c \in \mathcal{C}$, we assume that the set of chess-types $T_1^c, \ldots, T_m^c$ lists the possible conceptual changes that a chess piece of type T_i^c may endure. A conceptual change can then been represented as a re-classification, e.g., $\mathsf{Queen_i}(\mathtt{wq}, \mathtt{t}) \wedge \mathsf{Queen_j}(\mathtt{wq}, \mathtt{t}{+}1)$.[7] By contrast, the misinterpretation 'destroys' the previously existing piece and creates a new one, i.e., no migrations from $T_i^{c_1}$ to $T_j^{c_2}$ (with $c_1 {\neq} c_2$) are possible.

Our view is formalized by means of the axioms (a8)-(a12).[8] At any time, a chess-piece is conceptualized in some way (a8) and this conceptualization is unique (a9), (a10). Chess-pieces cannot migrate from a class of concepts to a different one (a10), they cannot survive a change of class, e.g., $\mathsf{Queen_i}(c,t) \wedge \mathsf{Rook_j}(c,t{+}1)$ is inconsistent. We illustrate (a11) and (a12) by means of an example. Whenever a player realizes that the rules for queen have to be changed from the ones that characterize $\mathsf{Queen_i}$ to the ones that characterize $\mathsf{Queen_j}$, she applies the new view to every queen (a11). Moreover (a12) assures that a substitution of $\mathsf{Queen_i}$ with $\mathsf{Queen_j}$ at $t{+}1$ implies that $\mathsf{Queen_j}$ has no instances at t and that $\mathsf{Queen_i}$ has no instances at $t{+}1$.

a8 $\textsc{ch}(x) \wedge \mathsf{E}(x,t) \rightarrow \bigvee_{i,c} T_i^c(x,t)$

a9 $T_i^c(x,t) \rightarrow \bigwedge_{j \neq i} \neg T_j^c(x,t)$

a10 $T_i^{c_1}(x,t) \rightarrow \forall t' (\bigwedge_{j, c_2 \neq c_1} \neg T_j^{c_2}(x,t'))$

a11 $T_i^c(x,t) \wedge T_i^c(y,t) \wedge T_j^c(x,t{+}1) \wedge \mathsf{E}(y,t{+}1) \rightarrow T_j^c(y,t{+}1)$

a12 $T_i^c(x,t) \wedge T_j^c(x,t{+}1) \rightarrow \neg \exists y (T_i^c(y,t{+}1)) \wedge \neg \exists y (T_j^c(y,t))$ (for $i \neq j$)

Note that we approached the distinction between conceptual changes and misinterpretation by syntactically fixing what migrations in type a piece can survive.[9] We did not provide a complete characterization of the range of possible changes of a chess-piece, because it seems to require modal reasoning and it would commit us to define what are the essential properties of a chess-piece. The only constraint we put is of a semantical nature: we exclude that an agent who knows that queens and rooks are different things may consistently believe that something that is a queen follows the rule of a rook. That is the meaning of our constraints on classes of chess-types.

As in the case of individual constants, to be as general as possible, we assume that all the types are local to players. We note $T_i^{\mathtt{W}}$ ($T_i^{\mathtt{B}}$) the ones in the report Π_{W} (Π_{B}). In addition, we assume that all the predicates in the same class have non-equivalent necessary conditions. This constraint does not hold for predicates in different classes. For instance, at a given time White can adopt as rules for queens the rules he adopted for rooks in the past.

[7] To simplify the notation we write $\mathsf{Queen_i}$ instead of $\mathsf{T_i^{queen}}$.

[8] Note that our definition does not exclude that two different chess-pieces are associated with the same set of predicates: for instance, we can assume that every rook is associated with the same list of possible types.

[9] This allows to group predicates *a posteriori* by looking at the migrations in a player's report.

2.1 Examples

We discuss a number of simple examples of a somehow queer chess game, where players may change their mind about chess-pieces and about the rules that regulate chess-types. These examples allow us to clarify what does it means to 'share' the game level.

Example 1. (*Rook or Queen*) Consider the three snapshots in Figure 1 represented by (f1), i.e., first, White moves the ♖ to b2, then Black moves ♜ in a2. Because the factual physical knowledge is shared among the players (f1) $\in \Pi_{\mathrm{W}}$ and (f1) $\in \Pi_{\mathrm{B}}$.

This sequence of moves is compatible with a number of scenarios in each player's perspective on the game-level. Firstly, we may assume a standard plain case: players coordinate and keep on playing because they share the same rules and they agree on what chess pieces exist and on their manifestations. For instance, there are no revisions, we have $\mathsf{Rook}^{\mathtt{W}}(\mathtt{br}_{\mathtt{W}},[\mathtt{t}_0,\mathtt{t}_2])$, $\mathsf{Rook}^{\mathtt{B}}(\mathtt{br}_{\mathtt{B}},[\mathtt{t}_0,\mathtt{t}_2])$, $\mathsf{M}($♜$, \mathtt{br}_{\mathtt{W}}, [\mathtt{t}_0,\mathtt{t}_2])$, $\mathsf{M}($♜$, \mathtt{br}_{\mathtt{B}}, [\mathtt{t}_0,\mathtt{t}_2])$,[10] and the necessary conditions of $\mathsf{Rook}^{\mathtt{W}}$ and $\mathsf{Rook}^{\mathtt{B}}$ are equivalent (they are the standard one for rooks).[11] In this case we say that the players agree on interpreting ♜ as rook. Similar for ♖, the players agree on interpreting ♖ as queen.

However, that is not the only possible scenario that allows the game to go on. For instance, suppose that, from $\mathtt{t}_0$ to $\mathtt{t}_2$, White is interpreting ♖ as rook instead of queen while Black is still interpreting ♖ as queen. Despite the disalignement, the two players still go on playing because the discrepancies in what pieces exist have not manifested in this phase of the game and could never manifest during a whole game, i.e., their conceptual alignment is indeterminate.

Example 2. (*A stone for a bit of wood*) Example 1 shows that the agreement on the interpretation of the physical level (and on the rules) is not necessary for keep on playing the game. We consider now an example where there are new physical entities that appear during the game: Black moves ♜ in b3, then White substitutes ♖ (originally in b1) with ● putting it c2 (see Figure 2). Suppose White is interpreting both ♖ and ● as queen, she changes the manifestation of the queen by using a stone, i.e., $\mathsf{M}($♖$, \mathtt{wq}_{\mathtt{W}}, [\mathtt{t}_0, \mathtt{t}_1])$, $\mathsf{M}($●$, \mathtt{wq}_{\mathtt{W}}, \mathtt{t}_2)$, and $\mathsf{Queen}^{\mathtt{W}}(\mathtt{wq}_{\mathtt{W}}, [\mathtt{t}_0, \mathtt{t}_2])$. At this point Black may wonder what have happened and may stop playing as she cannot understand which piece ● is the manifestation of (see (a7)). Or Black can assume a substitution, for instance agreeing with White on interpreting both ♖ and ● as queen.[12] Clearly, other scenarios are possible.

If it is true that rules create the very possibility of playing, as Searle states, it is also true that there is a non trivial sense according to which they are not enough in order to have the game. Whether or not is established by the rules that we can do something like changing a manifestation for another, it seems that we can anyway do the change without affecting the game-level. One could still follow the other rules that regulate

[10] We write $R(x_1, \ldots, x_n, [t_i, t_{i+n}])$ instead of $R(x_1, \ldots, x_n, t_i) \wedge \cdots \wedge R(x_1, \ldots, x_n, t_{i+n})$.

[11] Note that the necessary conditions are expressed only in term of the location L, a shared primitive. The equivalence of these conditions is then purely objective. The proof of the 'equivalence' of necessary conditions that involve private predicates is much more complex because it requires a link between all the involved private predicates.

[12] In the case of blindfold chess we could to say that a piece manifests itself through words. This is debated in the literature, for a discussion see [4].

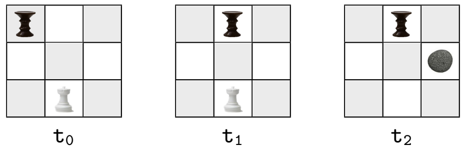

Fig. 2. Substituting a piece of wood with a stone.

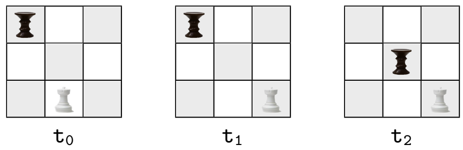

Fig. 3. Other three snapshots of the physical world.

the chess moves without any problem, even by maneuvering a stone instead of a bit of wood. This seems to reflect what one could call the twofold nature of institutional reality, in the sense that it seems to depend on the one hand on what is accepted by those who are dealing with it, but on the other also on what it is actually going on according to the rules, a reality emerging from both rules and practices.

Example 3. (*Rule Changing*) Example 1 is a trivial source of possible disagreement, as the standard moves of a queen include the moves of a rook. Consider now the scenario where White moves ♖ in c1, then Black moves ♜ in b2 (see Figure 3).

Again different game-reports are compatible with this physical sequence. Suppose that, at $\mathtt{t}_0$, both the players agree on interpreting ♜ as rook. At $\mathtt{t}_2$, Black moves by violating the standard rook-rules. Remember that we are assuming that the players are consistent with their rules and facts. This could mean that Black revised the rook-rules, i.e., $\mathtt{br}_\mathtt{B}$ persists but $\mathsf{Rook}^\mathtt{B}(\mathtt{br}_\mathtt{B}, \mathtt{t}_0)$ and $\mathsf{Rook}^\mathtt{B}_\mathsf{i}(\mathtt{br}_\mathtt{B}, \mathtt{t}_1)$.

Different scenarios are possible. White can refuse to keep on playing, because she maintains the standard rook-rule also at $\mathtt{t}_1$, i.e., $\mathsf{Rook}^\mathtt{W}(\mathtt{br}_\mathtt{W}, [\mathtt{t}_0, \mathtt{t}_1])$ and this fact together with the physical-facts are inconsistent with $\mathsf{M}($♜$, \mathtt{br}_\mathtt{W}, [\mathtt{t}_0,\mathtt{t}_2])$. In this case the game ends at $\mathtt{t}_2$ because White refuses the Black's move. Alternatively, White may accept the Black's move by revising her rook-rules to be consistent with this move. These new rules are compatible, but not necessarily identical, with Black's rules. The physical traces are not enough to completely determine the rules in Π_B and Π_W. One could also think that White changed her opinion about the existence of pieces: $\mathsf{M}($♜$, \mathtt{br}_\mathtt{W}, \mathtt{t}_0) \wedge \mathsf{M}($♜$, \mathtt{bk}_\mathtt{W}, [\mathtt{t}_1, \mathtt{t}_2])$—instead of $\mathsf{M}($♜$, \mathtt{br}_\mathtt{W}, [\mathtt{t}_0, \mathtt{t}_2])$—and $\mathsf{King}^\mathtt{W}(\mathtt{bk}_\mathtt{W}, [\mathtt{t}_1, \mathtt{t}_2])$.

This example shows that rules and pieces may keep on changing from time to time, as in the case of players learning how to play chess. Moreover, the change tells us that an actual game depends on moves and their acceptance, and this could be seen also as undermining the idea that rules are necessarily stable, they are given once and for all, as if, once the rule is in place, "all the steps are already taken", as Wittgenstein says ([14]: §219). Players may adapt, they may be open to change their rules to achieve a goal. It seems then that there should be an "agreement not only in definitions but also (...) in judgments" ([14]: §243). The actions and reactions of players are not simply important, they are, so to speak, the only thing that somebody has, for understanding what is going on when interacting with others. This is connected with the wittgensteinian idea of meaning as use: “the meaning of a word is its use in the language” ([14]: §43). At this point, one could say that the two players agree on the chess-pieces if they use them in the same way. But the previous examples show that we cannot be sure that we are using them in the same way by accessing only a limited sequence of configurations of the physical world. Wittgenstein says, “this was our paradox: no course of action could be determined by a rule, because every course of action can be made out to accord with the rule.” ([14]: §201). Different rules can be followed to achieve the same moves.

The acceptance of changes of rules during the game is debatable. According to many philosophers working on constitutive rules, if Black moves her rook diagonally, she is maybe playing some game, but not chess. Nonetheless, in our daily life, it becomes difficult to assess the exact boundaries of our institutional activities: changes and irregularities are quite common[13] and, as seen, to continue to play, White may accept the move of Black, revising her rules. Wittgenstein seems to be in this line of thought too:

> We can easily imagine people amusing themselves in a field by playing with a ball like this: starting various existing games, but playing several without finishing them, and in between throwing the ball aimlessly into the air, chasing one another with the ball, throwing it at one another for a joke, and so on. And now someone says: The whole time they are playing a ball-game and therefore are following definite rules at every throw. And is there not also the case where we play, and make up the rules as we go along? And even where we alter them — as we go along. ([14]: §83)

3 A potential illusion before the rules

Kripke in [5] interprets the aforementioned passage ([14]: §201) as the starting point to pose a paradox that undermines the very possibility of rule following. If we consider an individual in isolation the paradox in unsolvable, if we instead take into account agreement amongst members of a community, only a ‘sceptical’ solution is possible.

Kripke’s example focuses on rules governing the word ‘plus’. The symbol ‘$+$’ denotes the mathematical function of addition. Suppose that I have to make a computation, take ‘$68{+}57$’, I never performed before. In addition, suppose that my past computations never involved numbers greater than 56. I answer ‘125’, but a very bizarre sceptic asks me how I did it. The sceptic asks me to provide some proof that I did not change the interpretation of ‘plus’ and ‘$+$’, i.e., that in the past I did use ‘plus’ and ‘$+$’ to denote the addition and not, for example, to denote a function called ‘quus’ (symbolized by ‘$\oplus$’) defined by: $x \oplus y = x + y$, if $x, y < 57$; $x \oplus y = 5$, otherwise ([5]: 9). Clearly, following the rules for ‘plus’ is different from following the rules for ‘quus’, but my past answers (my ‘external behaviors’) are compatible with the possibility I was not adding but instead quadding. But the threat of the sceptic is even deeper, since it goes in arguing that even when mental facts are accepted as proof still the problem exists. Kripke argues that my mental states, my past intentions, my dispositions in my calculations, even an omniscient God that knows which continuation I was thinking of, do

[13] Ethnomethodological studies on board games seem to go also in this direction, see e.g. [8].

not help in founding a proof. If no adequate explanation is forthcoming—i.e., a rule is always compatible with different interpretations that lead to different results—then we must give up the idea that in the past I was following the rules for addition instead of the rules for quaddition. So there is no such thing as rule-following, no agreement or disagreement in accord with the rule. This goes against the inexorability of rules we mentioned in Example 3 and after Kripke has been called 'meaning determinism' [6], the idea that when one possesses a concept, "all future applications of it are determined (in the sense of being uniquely *justified* by the concept grasped)" ([5]: 107).[14]

In our framework, chess-pieces seem to play a role similar to symbols. Symbols may have multiple physical realizations: for instance the same word can be written or read aloud. Similarly, chess-pieces may have multiple manifestations, i.e., they abstract from the physical level to enter the game level.[15] In the Kripke example, symbols stand for functions defined by rules. Our chess-pieces are classified by chess-types that are given, but not completely defined, in terms of rules. Thus, we can simulate the change in the interpretation of a symbol as a re-classification of a chess-piece x, i.e., as $T^c_i(x,t) \wedge T^c_j(x,t')$ (with $i \neq j$). This re-classification implies a change in the rules the piece is submitted to, i.e., a change in the accepted moves. The computation required in the example of Kripke corresponds then to an acceptance of a move, i.e., instead of calculating the addition (quaddition) of two numbers, we check if the sequence of two configurations satisfy our constraints. To check the correctness of a move at a given time, to *compute* the result, one needs to follow some constraints or procedures. For instance, from an extensional perspective, to check if 125 is the result of $68 + 57$, it is enough to verify if the triple $\langle 68, 57, 125\rangle$ belongs to the function that is the interpretation of the symbol '$+$'. The set associated to '$+$' could vary in time (if we change the interpretation) but this does not mean that it is not possible, *at every instant*, to provide a justification of the answer (given at that time). Our reports go exactly in this direction, the classification $T^c_i(x,t)$ tells us what are the constraints x is submitted to, and these constraints can be used as justification. Clearly one observer could be in trouble in reconstructing the history of (the meaning of) '$+$' or finding the interpretation of '$+$' that is compatible with all the answers. But this is another problem.[16]

One objection to this dynamic view on rules is that for something to count as a rule, it has to norm homogeneously the behavior of a player throughout the game (this seems the position of Kripke). If this is so, the skeptical argument applies directly to our model as well. However, if we ask a player what is the rule for the queen that she was following during the game, a player may reply by saying that at any time she was changing the rules for the queen to cope with chess board situation. For instance, this shift appears explicitly in White's report as a switch of concepts, say from $\mathsf{Queen}^{\mathtt{W}}_{\mathtt{i}}$ to $\mathsf{Queen}^{\mathtt{W}}_{\mathtt{j}}$. That is, White was not following *a* rule for moving the queen, for example she was just trying to guess the right rule for the queen. But she may have no idea of what rule accounts

[14] Adding the deontic dimension complicates the picture without solving the problem.

[15] Note however that, at a given time, chess-pieces have a single manifestation, while symbols may have multiple realizations. In addition, we allow for ambiguous manifestations, manifestations that can be understood as different chess-pieces by different players.

[16] Similarly, arguments about the memory of what a player has done have only an epistemic impact.

for the full history of her moves. Kripke's sceptical argument is convincing if rules are statical and are perfectly accessible to players, on that, his examples are illuminating of this view: rules are exemplified by mathematical functions, that is something unambiguously defined and hardly subject to change. In case of social norms, such as those of chess game, the interpretation of a rule is challenging. Rules may be adapted to cope with new situations, take again Wittgenstein's example in §83: agents start various existing ball games never finishing them. According to Kripke, we would conclude that they are not following *a* rule. However, their behavior is not hectic and they are still coordinating and going along. Another way of viewing Wittgenstein's example is by saying that players agree on a rule that regulates how to change rules from time to time. If this is so, our chess player may reply to the question on what rule she was following by saying that she was following the rule that compel to change the rules for the queen in a given manner. Alternatively, one can look at the migrations of chess-pieces accepted by a player during the game. These migrations highlight (i) the acceptance of an originally inconsistent move, and (ii) how the player solved the inconsistency. Thus the 'rule' the player is following depends not only on the constraints provided by the chess-types, but also on her acknowledgment of the other's move.

The point is that skepticism about rule following strictly depends on what we view as a rule. Our model allows for both readings, as we can interpret the players' reports in two ways. Firstly, we can view a players' report as the history of his guessing of what is a chess rule: the player keeps on adapting her hypothesis to the new chess board situations. In this case, if the player is successful, at the end of the game, she may end up figuring out what is 'the' (or 'a') rule of the queen. On the other hand, at the end of the game, a player may have no idea of what is the rule of the queen. In any case, we cannot know whether the player was following any rule throughout the game, and scepticism applies. There is however another interpretation of the reports, that accounts for Wittgenstein's example of the ball game. Players can always play a queer game of chess that demands them to change the rules of the queen from time to time. The rule that the players are following is just the rule that compel them to change the rule of the queen, and that would be their answer to the question what rule were you following. Of course, analogous treatment applies to adding and quadding. In any case, according to both interpretations rules are dynamic. This is due to the fact that we are trying to take seriously, even if minimally, interaction into our model. Moreover, this is also in line with the skeptical, 'communitarian' solution posed by Kripke, according to which to follow a rule is to be under the scrutiny of others in a certain community: "if everyone agrees upon a certain answer, then no one will feel justified in calling the answer wrong" [5]. In our examples each player is ready to change her rules to continue the game, to align, since she knows that these changes are often necessary to interact.

What is then a rule? Far from providing a definition, we can say that, at every phase of the interaction, *even though* there is no shared rules or concepts—and we have to give up the epistemic privilege about social reality—still the interactants often behave as if there are. When we interact with others we could have no basis to know what rule they are following, or if they are following any rule at all. We try to find a way to fit a situation with others by reading off their behaviors, their actions and their reactions according to our behaviors. This means that if we read off the behavior and we predict a move that the player actually makes and the player has somehow accepted it, it could be possible that this reinforces in us a sense of following the same rule that allows us to keep on playing. We can have different rules—that objectively can be just personal constraints—and concepts, and pieces, and attributions but anyway keep going on playing. So it is not that the rule "traces the lines along which it is to be followed through the whole of space", but it is true that it is in this way, as Wittgenstein conclude ([14]: §219), that the rule "strikes me". Sameness of concepts is also potentially illusory. Nonetheless, the interactants can believe that the concepts are shared, since they are based on 'correct' (but just "up until now", before the other breaches them) predictions, and nothing prevents that we are just going on in the interaction by trials and errors, or, better, by acceptances and refusals.

References

1. Abbott, B.: Support for individual concepts. Linguistic and Philosophical Investigations 10, 23–44 (2011)
2. Brandom, R.: Between Saying and Doing. Towards an Analytical Pragmatism. OUP (2008)
3. Hindriks, F.: Constitutive rules, language, and ontology. Erkenntnis 71(2), 253–275 (2009)
4. Johansson, I.: Money and fictions. In: Kapten Mnemos Kolumbarium (2005)
5. Kripke, S.: Wittgenstein on Rules and Private Language. Basil Blackwell, Oxford (1982)
6. Kusch, M.: A Sceptical Guide to Meaning and Rules. Acumen, Chesham (2006)
7. Laurence, S., Margolis, E.: Concepts and cognitive science. In: Concepts: core readings, pp. 3–81. MIT Press Cambridge, MA (1999)
8. Liberman, K.: More Studies in Ethnomethodology. SUNY (2013)
9. McCarthy, J.: First order theories of individual concepts and propositions. Machine Intelligence 9 (1979)
10. Searle, J.R.: The Construction of Social Reality. The Free Press, New York (1995)
11. Thomasson, A.: Realism and human kinds. Philosophy and Phenomenological Research 68, 580–609 (2003)
12. Tuomela, R.: The Importance of Us. SUP, Stanford (1995)
13. Weber, M.: R[udolf] Stammler's "overcoming" of the materialist conception of history. In: Bruun, H.H., Whimster, S. (eds.) Max Weber: Collected Methodological Writings. Routledge (1907/2012)
14. Wittgenstein, L.: Philosophical Investigations. Basil Blackwell, Oxford, 4th edn. (1953/2009)