A measure theoretic paradox from a continuous colouring rule

Robert Simon[1] · Grzegorz Tomkowicz[2]

Abstract

Given a probability space $(X, \mathcal{B}, m)$, measure preserving transformations $g_1, \ldots, g_k$ of X, and a colour set C, a colouring rule is a way to colour the space with C such that the colours allowed for a point x are determined by that point's location in X and the colours of the finitely many $g_1(x), \ldots, g_k(x)$ (called descendants). We represent a colouring rule as a correspondence F defined on $X \times C^k$ with values in C. A function $f : X \to C$ satisfies the rule at x if $f(x) \in F(x, f(g_1x), \ldots, f(g_kx))$. A colouring rule is paradoxical if it can be satisfied in some way almost everywhere with respect to m, but not in **any** way that is measurable with respect to a finitely additive measure that extends the probability measure m defined on $\mathcal{B}$ and for which the finitely many transformations $g_1, \ldots, g_k$ remain measure preserving. We show that a colouring rule can be paradoxical when the $g_1, \ldots, g_k$ are members of a semi-group G, the probability space X and the colour set C are compact sets, C is convex and finite dimensional, and the colouring rule says if $c : X \to C$ is the colouring function then the colour $c(x)$ must lie (m a.e.) in $F(x, c(g_1(x)), \ldots, c(g_k(x)))$ for a non-empty upper-semi-continuous convex-valued correspondence F. Furthermore we show that this colouring rule has a stability property—there is a positive ϵ small enough so that if the expected deviation from the rule does not exceed ϵ then the colouring cannot be measurable in the same finitely additive way. As a consequence, there is a two-person Bayesian game with equilibria, but all ϵ-equilibria for small enough ϵ are not

✉ Robert Simon
  r.s.simon@lse.ac.uk

  Grzegorz Tomkowicz
  gtomko@vp.pl

1 Department of Mathematics, London School of Economics,
   Houghton Street, London WC2A 2AE, UK

2 Centrum Edukacji G2, ul. Moniuszki 9, 41-902 Bytom, Poland

\underline{\textcircled{\tiny ?}} Springer

measurable according to any finitely additive measure that respects the information structure of the game.

1 Introduction

In [7], we introduced colouring rules. We demonstrated several paradoxical colouring rules and proved that if there are finitely many colour classes and the measure preserving transformations belong to a group then any colouring of a paradoxical colouring rule has colour classes that jointly, with the measure preserving transformations and the Borel sets, define a *measurably G-paradoxical decomposition* (for the group G generated by the measure preserving transformations). We mean by this the existence of two measurable sets of different measures that are G-equidecomposable (see [7, Thm. 1]).

In the conclusion of [7], we asked whether a colouring rule could be paradoxical if the colour classes belonged to a finite dimensional convex set and the colouring rule was defined by an upper-semi-continuous convex-valued non-empty correspondence, as described above. We call such colouring rules *probabilistic* colouring rules. This means, among other things, that the choosing of colours could be according to a maximisation or minimisation of a continuous and affine evaluation of options, with indifference between two options implying indifference between all of their convex combinations.

Our main inspiration is the question whether measure theoretic paradoxes, such as the Banach Tarski Paradox, have any applications to areas beyond mathematics, such as physics or economics (see [8] for more details). A colouring rule could represent natural forces, and the lack of any measurable solution could represent a radical inability to predict their behaviour. A particular inspiration is the widely held belief in economic theory that although one cannot always accomplish optimisation goals through behaviour that is measurable with respect to a countably additive measure, one can do so with some finitely additive option. The problem with this belief is that there may be knowledge structures to the optimisation that cannot be altered when extending to a finitely additive measure. If those knowledge structures are defined through the use of ergodic operators, measure invariance of those ergodic operators may be required. Among other motivations, we wanted to show that there are some Bayesian games with equilibria but without measurable ϵ-equilibria for small enough positive ϵ according to any reasonable understanding of measurable.

The Kakutani Fixed Point Theorem is relevant to probabilistic colouring rules. If X is compact and the correspondence is independent of the location of x, meaning that the correspondence F is defined entirely on $g_1(x), \ldots, g_k(x)$, then the fixed point theorem shows there exists a constant colouring function (a fixed distribution) satisfying the colouring rule, hence it cannot be paradoxical. This is done by mapping a colour c in C to the k copies of C, namely $(c, c, \ldots c) \in C^k$, and then following the colouring rule back down to C via the correspondence F. Therefore we have to consider colouring rules that are dependent on the location in the space.

In Simon and Tomkowicz [6] we demonstrated a probabilistic colouring rule with a one dimensional continuum of colours such that after the correspondence is approxi-

mated by any small enough positive ϵ the colouring rule still had no Borel measurable solution. It was formulated as a Bayesian game for which there are equilibria, but a positive ϵ such that no ϵ-equilibrium is Borel measurable. In this paper, we demonstrate a Bayesian game whose information structure is defined by the action of a semi-group G such that the game does have equilibria, but there is a positive ϵ such that there is no finitely additive G-invariant measure μ extending the Borel measure together with an ϵ-equilibrium that is μ measurable.

In the next section we describe the probabilistic colouring rule and show that it is paradoxical. In Sect. 3 we look at approximating the colouring rule. In Sect. 4 we discuss Bayesian games and explain how they motivated us. We translate our paradoxical colouring rule to a two player Bayesian game and show how it does not have approximate equilibria that are measurable with respect to any finitely additive measure that respects the information structure of the game. In conclusion we consider related problems.

2 A probabilistic paradoxical colouring rule

Let T_1 and T_2 be two non-invertible generators (each generating a semi-group isomorphic to $\mathbf{N}$). We assume that $T_1^3 T_2 = T_2 T_1^3$ and there are no other relations. Let G be the cancellative semi-group with identity generated by T_1 and T_2. Let X be the set $\{0, 1\}^G$. We extend X to $X' = X \times \{a, b, c\}$ and let the symmetric group S_3 act on the three elements $\{a, b, c\}$. We assume that S_3 commutes with G and define G' to be the semi group $G \times S_3$. We need to include S_3 for the definition of the colouring rule. However it is mostly the colouring of $X \times \{a\}$ that matters.

For any $x \in X$ and $g \in G$, x^g stands for the g coordinate in x. With e the identity in G, the e coordinate of x is x^e. There is a canonical right semi-group action on X, namely $g(x)^h = x^{gh}$ for every $g, h \in G$. We use the canonical product topology on X. For every cylinder determined by particular choices of $\{0, 1\}$ we assign the probability $(\frac{1}{2})^k$ where k is the number of those choices determining the cylinder. With the Borel probability measure the semi-group G acts measure preserving on X (meaning that the measure of $T^{-1}(A)$ is equal to the measure of A, for any Borel set A and T in G). Any semi-group element acts measure preserving on any cylinder, due to the cancellation law, and this can be extended to any Borel set through approximation via cylinders. Likewise we give X' the Borel probability measure where each element in $\{a, b, c\}$ is given equal probability when paired with a Borel set in X. The probability of $A \times \{a, b, c\}$ in X' is given the same probability at that given to A in X, In this way G' acts measure preserving on X'. With m the canonical Borel measure of X, let m' be its extension to X'.

Where $g \neq h$ implies that $gx \neq hx$ is a Borel subset of measure one. This follows from the fact that the semi-group is countable. Without loss of generality, we will be interested only in this subset, and we ignore the set of measure 0 where this doesn't hold.

The set of colours C is $\Delta(\{1, 2, 3\}) := \{p = (p_1, p_2, p_3) \mid \forall i \ p_i \geq 0, p_1 + p_2 + p_3 = 1\}$ where δ_i is perceived to be all weight to the colour c_i. We represent the extremal colours c_i modulo 3, with $i = 1, 2, 3$ rather than 0, 1, 2.

Now we show how to colour each $x \in X'$ with a point in the simplex $\Delta(\{1, 2, 3\})$, with $p = (p_1, p_2, p_3) \in \Delta(\{1, 2, 3\})$ standing for the weights given to the three extremal colours. We colour according to the optimisation of a continuous function, and the resulting optimal solutions are represented by a correspondence F, defined on $X \times C^2$, where each $x \in X$ has two descendants $T_1 x$ and $T_2 x$. The continuous functions are defined by matrix multiplications.

For every $x \in X$ we define a matrix $A_x = \begin{pmatrix} 1 & r_{1,2}^x & r_{1,3}^x \\ r_{2,1}^x & 1 & r_{2,3}^x \\ r_{3,1}^x & r_{3,2}^x & 1 \end{pmatrix}$ such that $|r_{i,j}^x| \le$

$\frac{1}{100}$ for every choice of x and i, j. The $r_{i,j}^x$ are continuous functions of x in X, and are chosen such that for every $p \in \Delta(\{1, 2, 3\})$ the set of x such that two or three rows maximise $A_x p$ is a set of Borel measure zero. The following is one way to do that. There are six entries of the $r_{i,j}^x$ to determine. We place the elements of G into six infinite ordered collections. For each $l = 1, 2, \ldots$ and g_l the lth semi-group elements corresponding to $r_{i,j}^x$ we define $r_{i,j}^x := \frac{1}{100} \sum_{l=1}^{\infty} 2^{-l} x^{g_l}$. where the structure of X requires that $x^g \in \{0, 1\}$ for all $g \in G$.

We colour a point (x, b) by any convex combination of the c_i with the $i = 1, 2, 3$ whose rows maximise $A_{T_2 T_1 T_2(x)} p$, where p is the colour given to $(T_1 x, a)$.

We colour a point (x, c) by any convex combination of the c_i with the $i = 1, 2, 3$ whose rows maximise $A_{T_1 T_2 T_1(x)} p$, where p is the colour given to $(T_2 x, a)$.

The conceptionally complex part of the rule is how to colour a point (x, a).

We define two three dimensional matrices B_0 and B_1. The matrix B_0 is used when $x^e = 0$ and the matrix B_1 is used when $x^e = 1$. Both matrices have entries $(b_{i,j,k})$ only in $\{0, 1\}$ such that for each pair of columns j, k there is only one non-zero $b_{i,j,k}$ and it is equal to 1. There are two sets of columns and one row; the columns j and k correspond to the weights given to the colours c_j and c_k of (x, b) and (x, c) respectively.

A row $i = 1, 2, 3$ of B_1 or B_0 is evaluated in the following way. If $p \in \Delta(\{1, 2, 3\})$ is the colour given to (x, b) and $q \in \Delta(\{1, 2, 3\})$ is the colour given to (x, c) then the ith row is given the value $\sum_{j=1}^{3} \sum_{k=1}^{3} p_j q_k b_{i,j,k}$. The colouring rule requires that any convex combination of rows is chosen that maximise this row evaluation.

To define these matrices, we determine when the $b_{i,j,k}$ entry is 1. The first jth column represents the extremal colour c_j of (x, b) and the second kth column the extremal colour c_k of (x, c).

The matrix B_1, the case of $x^e = 1$, is easy to define. The $b_{i,j,k}$ entry is 1 if and only if $i = j + 1$.

The matrix B_0, the case of $x^e = 0$, is more complex.

If $k \ne 1$, and $j \ne 3$, then the $b_{i,j,k}$ entry is 1 if and only if $i = j$.

If $k \ne 1$, then the $b_{i,3,k}$ entry is 1 if and only if $i = 1$.

If $j \ne 3$, then the $b_{i,j,1}$ entry is 1 if and only if $i = j + 1$.

The $b_{i,3,1}$ entry is 1 if and only if $i = 3$.

This completes the definition of the correspondence F. We will also refer to F as the colouring rule. Notice that when $x^e = 0$ then the matrix B_0 has similarity to the conditions of the Hausdorff paradox, requiring that if the colour of x is an advancement by one on the colour of $T_1 x$ then the two points x and $T_2 x$ are coloured differently, one of these two points is coloured c_1 and the other takes a colour in $\{c_2, c_3\}$.

When (x, a), (x, b) or (x, c) gives all weight to an extremal colour, namely δ_{c_i} for some $i \in \{1, 2, 3\}$, then it is called *pure*.

A colouring $c : X' \to C$ *satisfies* the colouring rule F if the rule holds almost everywhere with respect to the probability distribution m'.

Lemma 1 *A colouring that satisfies the colouring rule F is pure almost everywhere with respect to the Borel measure m'.*

Proof There is no relation between $T_2 T_1 T_2$ and T_1, likewise between $T_1 T_2 T_1$ and T_2. Therefore for every choice of x the matrices at $(T_1^{-1}x, b)$ and at $(T_2^{-1}x, c)$ defining the colouring rule F are the matrices A_y for all $y \in X$. The conditional probability on those matrices is the same as the distribution on X. It follows by the construction of these matrices that regardless of the colour p at (x, a) the set of colours at $(T_1^{-1}x, b)$ and $(T_2^{-1}x, c)$ that are not pure with respect to m' is a subset of conditional measure 0 (here we identify the points in $T_i^{-1}x$ with the matrices A_y). By the definition of conditional probability, the conclusion follows for all (x, b) and (x, c). And therefore it follows for (x, a) also. □

Lemma 1 allows us to perceive a colouring of X' as being primarily a colouring of $X \times \{a\}$. With purity, the colours of (x, b) and (x, c) are merely conveying to (x, a) the colours of $(T_1 x, a)$ and $(T_2 x, a)$ in a way that those two colours are determining the colour of (x, a). From now on, by a colouring of X we mean a colouring of $X \times \{a\}$, where by the colour for x we mean the colour for (x, a). If a point is coloured purely with δ_i we will also write that it is coloured with c_i. Our main aim is to prove the following theorem:

Theorem 1 *For any finitely additive G-invariant measure μ on X' extending m' there exists no colouring $c : X' \to C$ that is μ measurable and satisfies the colouring rule F.*

Definition Semi-group elements $g_1, g_2, \ldots, g_k \in G$ are called *independent* if there are no relations between them. Two points x, x' are called *twins* if $T_1 x = T_1 x'$ and $T_2 x = T_2 x'$, meaning that the differ only by $x^e \neq (x')^e$. A point $x \in X$ is coloured *randomly* if x, the twin of x, $T_1 x$, and $T_2 x$ are all pure and x and its twin are coloured differently (meaning that for the twins x, x' the colour of $T_1 x$ is advanced by one to define the colour of x where $x^e = 1$ and the colour of $T_1 x$ is not advanced to define the colour of x' where $(x')^e = 0$). A colouring that is measurable with respect to some G-invariant finitely additive measure μ is called *measurable*.

Lemma 2 *If there is a positive measure of points that are coloured randomly for a measurable colouring satisfying the colouring rule F, and $g_1, g_2, \ldots, g_k$ are independent, then the probability of $\{x \mid g_i x$ is coloured $c_{n_i}, i = 1, 2, \ldots, k\}$ is the product $\prod_{i=1}^{k} q_{n_i}$ where q_n is the probability of $\{x \mid x$ is coloured $c_n\}$ for $n = 1, 2, 3$.*

Proof First we show that if there is a positive probability of random colouring, then from the stochastic matrix representing the transition of the distribution of colours of x to the colouring of $T_1^{-1}x$ there is a unique eigenvector in $\Delta(\{1, 2, 3\})$ with the eigenvalue 1 and the other eigenspaces correspond to eigenvalues with norms less than

1. Since by measure invariance we can assume that the probability distribution of the colours is the same at x and $T_1^{-1}x$ for the collection of all $x \in X$, this implies that the distribution is determined by this eigenvector.

Let p, q, r be half the probabilities for the random colouring conditioned on the colours c_1, c_2, c_3 respectively, with the half referring to the non-advancement of the colours which happens when $x^e = 0$. We assume that at least one of the p, q, r are positive and none are greater than $\frac{1}{2}$. The stochastic matrix in question is

$$\begin{pmatrix} p & 0 & 1-r \\ 1-p & q & 0 \\ 0 & 1-q & r \end{pmatrix}.$$

It has the characteristic polynomial $(p-x)(q-x)(r-x) + (1-r)(1-q)(1-p) = (1-r)(1-q)(1-p) + pqr - (qr+qp+pr)x + (p+q+r)x^2 - x^3$. The degree one polynomial $1-x$ divides this characteristic polynomial, leaving the second degree polynomial $1 - r - p - q + rp + rq + qp + (1 - r - p - q)x + x^2$ as a factor. Letting $b = 1 - r - p - q$ and $d = rp + rq + qp$ we have the roots $\frac{-b+\sqrt{b^2-4(b+d)}}{2}$ and $\frac{-b-\sqrt{b^2-4(b+d)}}{2}$. Assuming that $b^2 - 4(b+d)$ is not positive, the norm squared of these roots is equal to $b + d$, which is less than 1 because one of p, q, r must be positive and if, for example, $r > 0$ then $r > rp$. $b^2 - 4b$ cannot be positive if b is not negative. Since $b \geq -\frac{1}{2}$ and $\frac{1}{2} + \sqrt{2 + \frac{1}{4}} = 2$, the only way for the norm of $\frac{-b-\sqrt{b^2-4b}}{2}$ or $\frac{-b+\sqrt{b^2-4b}}{2}$ to reach 1 is if $b = -\frac{1}{2}$ and therefore $r = p = q = \frac{1}{2}$. But then the roots are really $\frac{-\frac{1}{2}+\sqrt{-\frac{3}{4}}}{2}$ and $\frac{-\frac{1}{2}-\sqrt{-\frac{3}{4}}}{2}$, with norms of $\frac{1}{2}$.

Due to the lack of any relation between the g_i, with respect to the Borel probability distribution whether the $(g_i x)^e$ are equal to 0 or 1 are independent choices over all the $x \in X$, hence also with any finitely additive measure extending the Borel measure. Assuming invariance of joint distributions of colour combinations before and after applying T_1^{-1}, using induction on the number of the g_i, and that the independence of the g_i implies the independence of the $T_1^n g_i$ for all positive n, we complete the proof with the following claim: □

Claim *Assume that there are two one-stage stochastic processes on two finite sets S and T respectively and a stochastic process defined on $S \times T$ such that the transitions are defined independently by transitions on S and T. Assume for each of the S and T processes that there is only one unique invariant distribution/eigenvector corresponding to an eigenvalue of norm 1. Furthermore assume that for each $s \in S$ and $t \in T$ that there is a positive probability that there is a transition to something other than s and t respectively. Then there is one unique invariant joint distribution on $S \times T$ defined by the independent distributions on S and T respectively.*

Proof of Claim: Let (s, t) be any pair of states in $S \times T$; we want to prove that any invariant probability for (s, t) is the same as ab where a is the invariant probability for s and b is the invariant probability for t. Let q be an invariant probability for (s, t); we want to show that $q = ab$. Let l_a be the probability, conditioned on the

state being at a, of leaving a on the next stage. Let l_b be the same probability for the state b. Let r_a be the probability of not being at a and returning to a on the next stage. Let r_b be the same probability for the state b. We can calculate q by $q = q - l_a q - l_b q + l_a l_b q + (1-a)r_a(1-l_b)b + (1-b)r_b(1-l_a)a + (1-b)r_b(1-a)r_a$. If these distributions on S and T are invariant independently, then $\bar{q} = ab$ is another solution for an invariant probability for (s, t). Therefore we can also write $ab = ab - l_a ab - l_b ab + l_a l_b ab + (1-a)r_a(1-l_b)b + (1-b)r_b(1-l_a)a + (1-b)r_b(1-a)r_a$, the same formula but with q replaced by ab. But then we can write $0 = q(l_a + l_b - l_a l_b) = ab(l_a + l_a - l_a l_b)$. With $l_a > 0$ and $l_b > 0$ we have $l_a + l_a - l_a l_b > 0$ and $q = ab$. q.e.d.

Notice that if $h = jg$ for semi-group elements g, h, j then the choice of gx for any $x \in X$ will determine the hx, including the value of $(hx)^e$, and therefore the $(gx)^e$ and $(hx)^e$ are dependent for the various choices of $x \in X$ in the maximal way that two variables can be dependent. If $kh = jg$ for some semi-group elements g, h, j, k, then starting at $kh(x) = jg(x)$ the dependence from kh and jg can have residual influences on the joint colour distribution at the set $\{(gx, hx) \mid x \in X\}$ before they reach their limit distributions. This residual dependence implies that the joint colour distribution at $\{(T_1 x, T_2 x) \mid x \in X\}$ should have some dependence coming from $T_1^3 T_2 = T_2 T_1^3$. However from a different perspective they should be independent! This contradiction drives our proof of Theorem 1.

Lemma 3 *There is no measurable colouring satisfying the colouring rule F such that there is a positive measure of points that are coloured randomly.*

Proof Let C_z be the stochastic matrix $C_z = \begin{pmatrix} \frac{1-z}{2} & 0 & \frac{2-z}{2} \\ \frac{z+1}{2} & \frac{1-z}{2} & 0 \\ 0 & \frac{z+1}{2} & \frac{z}{2} \end{pmatrix}$. Given an independent distribution of colours for the pair $T_1 x$ and $T_2 x$, the matrix determines the distribution of colours at x where z is the probability for the colour c_1 at $T_2 x$. Thus the entry c_{ij} of C_z determines the probability that x is coloured c_i given that $T_1 x$ is coloured c_j.

First consider the pair of independent elements $T_2 T_1$ and T_2^2 and let (y_1, y_2, y_3) be the global probabilities for the colours c_1, c_2, c_3, respectively. From Lemma 2 the joint distribution of the colours of $T_2 T_1 x$ and $T_2^2 x$ are that determined by the products of the y_i, meaning that the probability of $T_2 T_1 x$ and $T_2^2 x$ coloured c_i and c_j respectively are $y_i \cdot y_j$. But as the colouring is measurable, using that T_2 is measure preserving, we have to assume that the probability of the $T_1 x$ and $T_2 x$ coloured c_i and c_j respectively is also $y_i \cdot y_j$. Applying the matrix C_z with $z = y_1$ for $C_{y_1}(y_1, y_2, y_3)^t = (y_1, y_2, y_3)^t$ we get the following three equations with three variables:

$$y_1 = \frac{1}{2}(y_3) + \frac{1}{2}((1-y_1)y_1 + (1-y_1)y_3)$$

$$y_2 = \frac{1}{2}(y_1) + \frac{1}{2}((1-y_1)y_2 + y_1^2)$$

$$y_3 = \frac{1}{2}(y_2) + \frac{1}{2}(y_1 y_3 + y_1 y_2).$$

This solves to $y_1 = 3 - \sqrt{7}$, $y_2 = 3 - \sqrt{7}$, and $y_3 = 2\sqrt{7} - 5$. These are the global probabilities for the colours c_1, c_2, c_3 respectively. Approximately this is the triple $(.35425, .35425, .2915)$. The reason for a smaller probability for the colour c_3 results from the tendency to move away from this colour with greater probability than toward it.

Second consider the seven semi-group elements.

$$g_1 := T_2^4 T_1 T_2^3,$$
$$g_2 := T_2^4 T_1 T_2^2 T_1 T_2,$$
$$g_3 := T_2^4 T_1 T_2^2 T_1^2 T_2,$$
$$g_4 := T_2^4 T_1 T_2^2 T_1^3,$$
$$g_5 := T_2^4 T_1 T_2 T_1^4,$$
$$g_6 := T_2^4 T_1 T_2^2 T_1^2,$$
$$g_7 := T_2^4 T_1 T_2^2 T_1.$$

We show that they are independent, meaning that there is no relations between them. Make the revision that T_1 is invertible and T_1^3 is the identity. If the resulting elements still have no relations between them, there was no relation before this revision. We can now proceed with the assumption that there are no relations other than $T_1^3 = e$.

We have the reductions

$$h_1 := T_2^4 T_1 T_2^3,$$
$$h_2 := T_2^4 T_1 T_2^2 T_1 T_2,$$
$$h_3 := T_2^4 T_1 T_2^2 T_1^2 T_2,$$
$$h_4 := T_2^4 T_1 T_2^2,$$
$$h_5 := T_2^4 T_1 T_2 T_1,$$
$$h_6 := T_2^4 T_1 T_2^2 T_1^2,$$
$$h_7 := T_2^4 T_1 T_2^2 T_1.$$

Notice that any combinations of the h_i are separated by T_2^4 or a higher power of T_2, so that we can identify when one element ends and the next begins. Removing these bookends of T_2^4, T_2^5, T_2^6, or T_2^7, we are left with the interiors T_1, $T_1 T_2^2 T_1$, $T_1 T_2^2 T_1^2$, $T_1 T_2 T_1$. All but the $T_1 T_2 T_1$ (from h_5) are repeated twice. The two uses of T_1 are distinguished by the powers of T_2 succeeding them (T_2^7 vs T_2^6). Likewise the two uses of $T_1 T_2^2 T_1$ can be distinguished by the power of T_2 succeeding them (T_2^5 vs T_2^4) and the same is true of $T_1 T_2^2 T_1^2$ (T_2^5 vs T_2^4). It follows that all seven elements $g_1, \ldots, g_7$ also must be independent.

By Lemma 2 the probability of $\{x \mid g_i x$ coloured $c_{n_i}\}$ for any choice of $n_1, \ldots, n_7$ is equal to the product $\prod_{i=1}^{7} y_{n_i}$. Due to the measure preserving property of $T_2^4 T_1 T_2$, we can assume that the probability of $\{x \mid g_i' x$ coloured $c_{n_i}\}$ for any choice of $n_1, \ldots, n_7$ is equal to the product $\prod_{i=1}^{7} y_{n_i}$ for the elements

$$g_1' := T_2^2,$$
$$g_2' := T_2 T_1 T_2,$$
$$g_3' := T_2 T_1^2 T_2,$$
$$g_4' := T_2 T_1^3,$$
$$g_5' := T_1^4,$$
$$g_6' := T_2 T_1^2,$$
$$g_7' := T_2 T_1.$$

Now consider the defining condition that $g_4' x = T_1^3 T_2 x = T_2 T_1^3 x$ for all x. For each of the three colour choices at $T_1^3 T_2 x$ we determine the distribution on the colours of x by building up to a joint distribution for $T_1 x$, $T_2 x$, and x through the two pathways, the left pathway going through three applications of T_1^{-1} and the right pathway going first through T_2^{-1} followed by two applications of T_1^{-1}. Due to the above mutual independence of colour distributions of the $g_i' x$ and their locations in relation to the two pathways we can use the matrices C_z to determine these distributions, including the last step of determining the colour distribution of x (as pointed out at the beginning of the proof), conditioned on the colour of $T_1^3 T_2 x$. There are three matrices most relevant to our calculations, C_1, C_0, and C_{y_1}.

The distribution of colours for $T_2 x$, conditioned on a choice of colour for $T_1^3 T_2(x)$ is determined by

$$C_{y_1}^3 = \begin{pmatrix} \frac{1}{8}(50\sqrt{7} - 129) & \frac{1}{4}(23 - 8\sqrt{7}) & \frac{1}{8}(19\sqrt{7} - 49) \\ \frac{1}{8}(216 - 81\sqrt{7}) & \frac{1}{8}(50\sqrt{7} - 129) & \frac{1}{4}(23 - 8\sqrt{7}) \\ \frac{1}{8}(31\sqrt{7} - 79) & \frac{1}{8}(91 - 34\sqrt{7}) & \frac{1}{8}(11 - 3\sqrt{7}) \end{pmatrix}.$$

To determine the distribution of colours for $T_1 x$ we need to calculate $C_{y_1}^2 C_0(y_1, y_2, y_3)$ and $C_{y_1}^2 C_1(y_1, y_2, y_3)$, the former for what happens when $T_1^3 T_2 x$ is coloured c_2 or c_3 and the latter for $T_1^3 T_2 x$ coloured c_1. For the former we get

$$\begin{pmatrix} \frac{1}{4}(11 - 4\sqrt{7}) & \frac{1}{4}(5\sqrt{7} - 11) & \frac{1}{4}(\sqrt{7} - 1) \\ \frac{1}{2}(6\sqrt{7} - 15) & \frac{1}{4}(11 - 4\sqrt{7}) & \frac{1}{4}(5\sqrt{7} - 11) \\ \frac{1}{4}(23 - 8\sqrt{7}) & \frac{1}{4}(4 - \sqrt{7}) & \frac{1}{2}(8 - 3\sqrt{7}) \end{pmatrix} \begin{pmatrix} \frac{3}{2}\sqrt{7} - \frac{7}{2} \\ 3 - \sqrt{7} \\ \frac{3}{2} - \frac{\sqrt{7}}{2} \end{pmatrix}$$
$$= \left(\frac{1}{8}(117\sqrt{7} - 307), \frac{1}{4}(258 - 97\sqrt{7}), \frac{1}{8}(77\sqrt{7} - 201) \right).$$

For the latter we get

$$
\begin{pmatrix}
\frac{1}{4}(11-4\sqrt{7}) & \frac{1}{4}(5\sqrt{7}-11) & \frac{1}{4}(\sqrt{7}-1) \\
\frac{1}{2}(6\sqrt{7}-15) & \frac{1}{4}(11-4\sqrt{7}) & \frac{1}{4}(5\sqrt{7}-11) \\
\frac{1}{4}(23-8\sqrt{7}) & \frac{1}{4}(4-\sqrt{7}) & \frac{1}{2}(8-3\sqrt{7})
\end{pmatrix}
\begin{pmatrix}
\sqrt{7}-\frac{5}{2} \\
3-\sqrt{7} \\
\frac{1}{2}
\end{pmatrix}
$$
$$
= \left(\frac{1}{8}(95\sqrt{7}-248), \frac{1}{8}(429-161\sqrt{7}), \frac{1}{8}(66\sqrt{7}-173) \right).
$$

For each of the three possibilities, $T_1^3 T_2 x = T_2 T_1^3 x$ coloured c_1, c_2, or c_3, we do the calculations for the three colours at $T_1^3 T_2 x$ separately and recombine the results, using the law of total probability, according to (y_1, y_2, y_3).

For $T_1^3 T_2 x$ coloured c_1, we have to calculate $C_{\frac{1}{8}(50\sqrt{7}-129)}(\frac{1}{8}(95\sqrt{7}-248), \frac{1}{8}(429-161\sqrt{7}), \frac{1}{8}(66\sqrt{7}-173)^t =$

$$
\begin{pmatrix}
\frac{1}{16}(137-50\sqrt{7}) & 0 & \frac{1}{16}(145-50\sqrt{7}) \\
\frac{1}{16}(50\sqrt{7}-121) & \frac{1}{16}(137-50\sqrt{7}) & 0 \\
0 & \frac{1}{16}(50\sqrt{7}-121) & \frac{1}{16}(50\sqrt{7}-129)
\end{pmatrix}
\begin{pmatrix}
\frac{1}{8}(95\sqrt{7}-248) \\
\frac{1}{8}(429-161\sqrt{7}) \\
\frac{1}{8}(66\sqrt{7}-173)
\end{pmatrix}
$$
$$
= \left(\frac{1}{128}(43635\sqrt{7}-115411), \frac{1}{128}(178381-67402\sqrt{7}), \frac{1}{128}(23767\sqrt{7}-62842) \right)
$$

For $T_1^3 T_2 x$ coloured c_2, we have to calculate $C_{\frac{1}{4}(23-8\sqrt{7})}(\frac{1}{8}(117\sqrt{7}-307), \frac{1}{4}(258-97\sqrt{7}), \frac{1}{8}(77\sqrt{7}-201)^t =$

$$
\begin{pmatrix}
\frac{1}{8}(8\sqrt{7}-19) & 0 & \frac{1}{8}(8\sqrt{7}-15) \\
\frac{1}{8}(27-8\sqrt{7}) & \frac{1}{8}(8\sqrt{7}-19) & 0 \\
0 & \frac{1}{8}(27-8\sqrt{7}) & \frac{1}{8}(23-8\sqrt{7})
\end{pmatrix}
\begin{pmatrix}
\frac{1}{8}(117\sqrt{7}-307) \\
\frac{1}{4}(258-97\sqrt{7}) \\
\frac{1}{8}(77\sqrt{7}-201)
\end{pmatrix}
$$
$$
= \left(\frac{1}{32}(9856-3721\sqrt{7}), \frac{1}{64}(13429\sqrt{7}-35509), \frac{1}{64}(15861-5987\sqrt{7}) \right).
$$

For $T_1^3 T_2 x$ coloured c_3, we have to calculate $C_{\frac{1}{8}(19\sqrt{7}-49)}(\frac{1}{8}(117\sqrt{7}-307), \frac{1}{4}(258-97\sqrt{7}), \frac{1}{8}(77\sqrt{7}-201)^t =$

$$
\begin{pmatrix}
\frac{1}{16}(57-19\sqrt{7}) & 0 & \frac{1}{16}(65-19\sqrt{7}) \\
\frac{1}{16}(19\sqrt{7}-41) & \frac{1}{16}(57-19\sqrt{7}) & 0 \\
0 & \frac{1}{16}(19\sqrt{7}-41) & \frac{1}{16}(19\sqrt{7}-49)
\end{pmatrix}
\begin{pmatrix}
\frac{1}{8}(117\sqrt{7}-307) \\
\frac{1}{4}(258-97\sqrt{7}) \\
\frac{1}{8}(77\sqrt{7}-201)
\end{pmatrix}
$$
$$
= \left(\frac{1}{64}(10663\sqrt{7}-28183), \frac{1}{64}(41681-15746\sqrt{7}), \frac{1}{64}(5083\sqrt{7}-13434) \right).
$$

Combining these three vectors according to the distribution $(3-\sqrt{7}, 3-\sqrt{7}, 2\sqrt{7}-5)$ of the three colours at $T_1^3 T_2 x$ we get a different distribution for the colours at x,

namely $(\frac{1}{64}(75588 - 28561\sqrt{7}), \frac{1}{128}(95189\sqrt{7} - 251801), \frac{1}{128}(100753 - 38067\sqrt{7}))$, a contradiction. In decimals this corresponds approximately to $(.35464, .35486, .29050)$, (different from the original $(.35425, .35425, .2915) = (3 - \sqrt{7}, 3 - \sqrt{7}, 2\sqrt{7} - 5))$. $\qquad\square$

The source of the discrepancy in the final distribution at the top comes from different starting distributions at $T_2 T_1^3 x$ and $T_1^3 T_2 x$. If there was no commuting of T_1^3 and T_2, there would be two different points and the probability of a start of c_i at $T_1^3 T_2 x$ and c_j at $T_2 T_1^3 x$ would be $y_i y_j$ with $(y_1, y_2, y_3) = (3 - \sqrt{7}, 3 - \sqrt{7}, 2\sqrt{7} - 5)$. But with the commuting there are no mixed starts.

Finally with the above lemmas, we can prove Theorem 1.

Proof of Theorem 1 By the three lemmata, a measurable colouring satisfying the colouring rule must have no positive subset of points coloured randomly. That means, with any probability approaching 1, that the colours cycle through with x coloured c_{i+1} whenever T_1 is coloured c_i. So with the measure preserving application of T_1 we see that $\frac{1}{3}$ of the space is coloured c_1. On the other hand, by T_2 being measure preserving, as with the Hausdorff paradox, this implies that the probability of the space coloured c_1 must be arbitrarily close to $\frac{1}{2}$, a contradiction. $\qquad\square$

There is of course a non-measurable colouring of the space. Start at some x. Let the colours cycle through c_i by the repeated application of T_1^{-1}, increasing the colour by one, and T_1, decreasing the colour by one. Choose some y already coloured this way and colour $T_2 y$ so that that $T_2 y$ is coloured c_1 if and only if y is not coloured c_1. Extend this choice for $T_2 y$ with the repeated application of T_1^{-1} and T_1 to $T_2 y$. Notice that the commuting of T_1^3 and T_2 does not get in the way of this pattern. This can be continued, but there is a general solution to the existence of a non-measurable colouring which includes this type of colouring. This general solution is presented in the next theorem. It demonstrates the other half of the argument that F is a paradoxical colouring rule, that the colouring rule F can be satisfied.

Let X^* be the x in X where $g \neq h$ implies that $gx \neq hx$.

Theorem 2 *There are pure colourings of X^* that satisfy the colouring rule F.*

Proof Let a subset A of X be called *closed* if whenever $T_1(x) \in A$ and $T_2(x) \in A$ then also x is in A. Let $\overline{A}$ stand for the closure of A. A subset A of X is called *pyramidic* if whenever $x \in A$ then the semi-group orbit Gx is in A. Notice that the closure of a pyramidic set is pyramidic.

For any pyramidic set B and any pure colouring of B consistent with the colouring rule there is a deterministic way to colour the closure $\overline{B}$ according to the colouring rule. We define a partially ordered set on the pairs (B, c) where B is a pyramidic and closed set and c is a colouring of B according to the rule. We say that $(B, c) \geq (B', c')$ if B contains B' and c restricted to B' is c'.

With Zorn's lemma there is a maximal element to any tower of the partial order. We observe that it is not possible for a closed and pyramidic B to be maximal and also that there is some $x \in X^*$ such that $x \notin B$. So it is enough to show that one can always extend the colouring form (B', c') to some (B, c) in a consistent way.

Consider the orbit Gx and the sequence A_l where A_l is the subset of Gx such that $y \in A_l$ if and only if $y = gx$ with g a word of length l and $y \notin B$. In any way we colour A_l, the colouring rule extends to a colouring of the closure of $A_l \cup B$ that follows the rule and does not change the colours of B. Indeed, let f_l be a sequence of such colourings of $A_l \cup B$. As Gx is a countable set, we can observe that a subsequence of the f_l converges point-wise to a colouring function f on $Gx \cup B$. We then extend this function f to a pure colouring on the closure of $Gx \cup B$. □

3 Optimality and stability

The colouring rule F is already formulated as a problem of local optimisation according to an objective function. We can relax the rules, so that for some given $\delta > 0$ it is required that the colours chosen at all points are within δ of optimality. We call this *pointwise δ-optimality*. But that is only one concept of approximate optimality, that at each individual point there is no gain by more than δ through a different choice of colour. We seek a broader concept. For each $x \in X$ let $t(x)$ be the possible improvement in the objective function at x, keeping the colouring for all other $y \neq x$ fixed. Let μ be a G-invariant finitely additive extension. A colouring is *γ-stable* if the μ-expectation of $t(x)$ is no more than $\gamma \geq 0$, meaning that there is no finite disjoint collection $A_1, \ldots, A_n$ of μ measurable sets such that the objection function can be improved by at least t_i at all points in A_i and $\sum_{i=1} \mu(A_i)t_i$ is greater than γ.

Another way of understanding γ-stability is that X is an uncountable space of human society or molecules, and the solution is γ-stable if the gains from the individual deviations do not add up to an expectation of γ.

There are two ways that a measurable colouring must obey γ-stability. First, the set where there is significant divergence from optimality must be small. Second, where divergence from optimality exists in a subset of large measure, that divergence must be small (the first concept we presented). That can be formalised in the following way: if a colouring is $\epsilon \cdot \delta$-stable, then the subset where it diverges from optimality by more than δ cannot be of measure more than ϵ.

By the continuity of Borel measure, with $p \in \Delta(\{1, 2, 3\})$ fixed and as δ goes down to 0 the set of x where two rows of A_x have expectations within δ of the optimal choice falls in Borel measure to 0. As the finitely additive measure must be an extension of the Borel measure, the same is true for the finitely additive measure. This allows the following corollary, whose proof is a refinement of the proof of Theorem 1.

Corollary 1 *For small enough γ there is no finitely additive G-invariant measure extending the Borel measure with a γ-stable measurable colouring for the colouring rule F, and hence there is a paradoxical colouring rule defined by a continuous function.*

Proof Let $\delta > 0$ be fixed, and consider the subset X_δ of X where only one row of A_x is optimal and the other two are not within δ of being optimal. As δ goes to zero the probability (according to m) of $X \backslash X_\delta$ goes to zero. Let $\rho(\delta)$ be a function of δ such that $\rho(\delta)$ goes to 0 as δ goes to 0 and $\rho(\delta)$ is greater than the probability of $X \backslash X_\delta$ for all positive δ.

We break the measurable colourings of X' into two cases, (1) those involving at least $\frac{1}{100}$ of the space coloured randomly and (2) those with less than $\frac{1}{100}$ of the space coloured randomly. For every $\epsilon > 0$ require that the probability for not having a pure colour or that the one extremal colour chosen is not optimal is less than ϵ. We have only to determine a positive ϵ and positive δ small enough so that $\rho(\delta) < \epsilon$ and ϵ is small enough to assure, for both cases, a discrepancy in the measure of the set where purity holds and c_1 is chosen. For the latter case (2) this is easy, $\epsilon < \frac{1}{1000}$ suffices for a contradiction (as the subset coloured c_1 would have to be simultaneously below $\frac{3}{8}$ and above $\frac{11}{24}$). For the former case (1), due to the continuity of the determinant, and therefore also the characteristic polynomial, following the argument in the proof of Lemma 2 the $\frac{1}{100}$ of the space using random generation implies the existence of a $d > 0$ and an $\bar{\epsilon} > 0$ such that the other eigenvalues (other than 1) in the transition matrix of the colours have norms less than $1 - d$ for all $0 < \epsilon < \bar{\epsilon}$.

Due to G-invariance of any proposed finitely additive measure, and assuming $\epsilon < \bar{\epsilon}$, starting at any distribution on colours and applying T_1^{-1} there is uniform minimal convergence rate to a unique invariant distribution, meaning that at each stage the difference between the distribution and the limit distribution is no more than $1 - d$ times what it was on the previous stage. As the transitions from the random process determined by $y^e = 0$ or $y^e = 1$ they will remain independent at the various $y = g_i x$ for independent elements $g_1, \ldots, g_k$ and they will dominate any distributional dependency from a small set of measure less than ϵ of not following the colouring rule F, we can repeat the arguments of Lemma 2, showing that the joint distributions of the $g_i' x$ for the seven g_i' are independent in the limit as ϵ goes to 0. The same can be done for the colour distribution on the whole space, that there is convergence to independent distributions of colours with T_1 and T_2, using the independence of $T_2 T_1$ and T_2^2 and measure invariance.

Due to the constant d and the existence of unique invariant distributions, as ϵ goes to 0 the convergence to equalities of the three equations with three variables y_1, y_2, y_3 from Lemma 3 implies that the unique invariant distributions on colours converges to the same fixed point distribution $(y_1, y_2, y_3) = (3 - \sqrt{7}, 3 - \sqrt{7}, 2\sqrt{7} - 5)$. Although the calculations by which one determines the distribution of colours at x from the point $T_1^3 T_2 x = T_2 T_1^3 x$ don't hold perfectly due to a subset of size up to ϵ where the colour rule does not hold, nevertheless with ϵ small enough they show a persistent $\frac{1}{5000}$ discrepancy in the probability given to the colour c_1. Having determined positive $\epsilon < \bar{\epsilon}$ and $\rho(\delta) < \epsilon$ small enough for both cases, we choose δ small enough to guarantee $\rho(\delta) < \epsilon$. The lack of $2\delta\epsilon$-stability follows.

Because the matrices A_x change continuously on the Cantor set X and the two matrices B_0 and B_1 are defined on disjoint clopen sets (hence together change continuously), and because the T_1 and T_2 are continuous functions, two points close to each other are close in terms of the consequences of colour choice. With the colouring rule F defined through the optimisation of a continuous function, optimised at the correspondence F, and with uniform continuity of the optimising functions (from X' and C compact), any sequence of colouring rule correspondences $F_1, F_2, \ldots$ that approximate the colouring rule F are also approximating its optimisation. With the correspondence F non-empty, upper-semi-continuous and convex valued, for every sequence of positive $\gamma_1, \gamma_2, \ldots$ converging to 0 there is a sequence of continuous func-

tions $f_i : X' \times C^2 \to C$ that approximate the correspondence F and the satisfaction of f_i implies point-wise γ_i optimality with respect to the correspondence F. By the above, there is a positive $\overline{\gamma}$ where satisfaction of f_i through measurable colouring is no longer possible when $\gamma_i < \overline{\gamma}$. Therefore we get eventually paradoxical colouring rules defined by continuous functions. □

We don't use the full force of γ-stability in showing that there are paradoxical colouring rules from continuous functions. However there is one application of γ-stability that does use that the probability of significant deviation from optimality is limited in probability.

4 Bayesian games

An important part of economic theory is the study of incomplete information. The idea is that some economic agent has some information that the others do not have, and this private information has to be used carefully to that player's advantage. It could be only one player with private information or it could be all the players. Often this situation can be modelled as a *Bayesian* game.

The connection to the above colouring rule F is that there is a Bayesian game played on the same probability space X' for which local optimising behaviour by a player is equivalent to satisfaction of the colouring rule F at an appropriate point and equilibrium behaviour is equivalent to satisfaction of the colouring rule F almost everywhere. Furthermore, a γ-equilibrium of the game is equivalent to a colouring with the γ-stability property.

Our interest in paradoxical colouring rules came originally from game theory, from the desire to show that **all**, not just some, equilibria of a game are not measurable. Simon [5] showed that there is a Bayesian game which had no Borel measurable equilibria, though it had non-measurable equilibria. The infinite dihedral group, an amenable group, acted on the equilibria in a way that prevented any equilibrium from being measurable. Simon and Tomkowicz [6] showed that there is a Bayesian game with non-measurable equilibria but no Borel measurable ϵ-equilibrium for small enough positive ϵ. That construction involved the action of a non-amenable semi-group.

A few words are necessary concerning the way a Bayesian game is played. There is a probability space $(\Omega, \mathcal{F}, m)$; nature chooses a point x in the space Ω according to the probability distribution m defined on a sigma algebra $\mathcal{F}$. There are two approaches to defining the information, strategies and payoffs of a player j.

In one approach, for each player j there is a sigma algebra $\mathcal{F}_j$ smaller than $\mathcal{F}$ such that the strategy of player j is a function measurable with respect to $\mathcal{F}_j$.

With the other approach we assume that a player j has a partition $\mathcal{P}_j$ of the space Ω—if nature chooses some $x \in \Omega$ the player j learns that nature's choice lies in the $B \in \mathcal{P}_j$ such that $x \in B$. If two points x, y belong to the same partition member B then player j cannot distinguish between x and y and must act identically at x and y.

The different approaches result in different ways to understand what is the strategy and payoff of a player.

With the measurable function approach the evaluation of a player's strategy is determined by the strategies of the other players doing the same, and because all their sigma algebras are contained in $\mathcal{F}$, the evaluation of the result goes through the probability distribution m. We call this the Harsanyi approach.

The partition approach we call the Bayesian approach. With the Bayesian approach each player j has a probability distribution on each set in $\mathcal{P}_j$ and evaluates its actions according to the actions of other players inside of the appropriate member of $\mathcal{P}_j$. Notice that the Bayesian approach doesn't really need a probability measure m for the whole space, though we include it because we want to link up the two approaches.

Though the two approaches are different, they can be related. If every member of $\mathcal{P}_j$ is in $\mathcal{F}$, we may move from the Bayesian approach to the Harsanyi approach. A sigma algebra $\mathcal{F}_j$ for player j is defined in the canonical way; a set $A \in \mathcal{F}$ is in $\mathcal{F}_j$ if and only for every set $B \in \mathcal{P}_j$, $A \cap B$ is either B or the empty set. If additionally the player j's probability distributions on each member of $\mathcal{P}_j$ form a regular conditional probability with respect to m and the $\mathcal{F}_j$ are generated as above, we complete the move to the Harsanyi approach.

Throughout we assume that the payoff of each player is affine with respect to changes in any one player's strategy, both with the Harsanyi measurable perspective and with the local Bayesian perspective. With the Harsanyi perspective one could say that the payoff functions are multi-linear in the potentially infinite dimensional space of measurable strategies; when there are finitely many players and the strategy spaces are finite dimensional the payoff for each player is represented by a multi-dimensional matrix. One can also consider more complex payoff structures, but even in the most trivial information structures (for example games of complete information) the existence of an equilibrium is not guaranteed when optimality doesn't occur in a convex set.

Both approaches to what defines a strategy and a payoff have their strengths and weaknesses. The Bayesian approach is more inclusive because it does not require that strategies are measurable. But in general, the Bayesian approach is more problematic. With the Bayesian approach, an evaluation of an action by player j in some $A \in \mathcal{P}_j$ may be impossible because within A the strategies of the other players may not be measurable with respect to the probability distribution player j has in the set A. But if each member in $\mathcal{P}_j$ is finite there is not a problem. Also if there is sufficient structure to the collection $\mathcal{P}_j$, and we have not yet determined the local probability distributions for each player, we can determine a probability distribution on each member B in $\mathcal{P}_j$ as a regular conditional probability with support on B (see [1]), and therefore add a link between the Hansanyi approach to the Bayesian approach.

When the Harsanyi and Bayesian approaches are linked by a regular conditional probability, the importance of measure preserving invariance to finitely additive extensions can be observed. A regular conditional probability on some $A \in \mathcal{P}_j$ must respect (almost everywhere) any measure preserving transformations taking place within the set A. By this we mean that if B is a measurable subset in $A \in \mathcal{P}_j$ such that $T^{-1}(B)$ is also contained in A, then (almost everywhere) the regular conditional probability for that player at that set must give B and $T^{-1}(B)$ the same measure. If a finitely additive measure on Ω that extends the original measure m doesn't respect the local probability distributions of the players, the game is distorted and the players' interests would be

no longer represented. If these local probability distributions are defined by measure preserving transformations then it makes sense that the finitely additive measure must keep those transformations measure preserving.

An ergodic game, (full definition in [5]), is one where the most important properties are that for each player j each member of $\mathcal{P}_j$ are finite and the player's local belief at each such member of $\mathcal{P}_j$ form a regular conditional probability. With ergodic games, neither the Bayesian approach nor the Harsanyi approach is problematic. Strictly speaking, the Bayesian game we present below is not ergodic because the partition members of $\mathcal{P}_j$ for one of the players are not finite. We define a *semi-ergodic* game to have all the same properties of an ergodic game with the relaxation that some members B of $\mathcal{P}_j$ may be infinite, however for every such infinite B there a finite subset $B' \subseteq B$ such that inside the set $B \backslash B'$ no player (including Player j) has influence over the payoff of Player j. A *quasi-ergodic* game has the further relaxation that inside the set $B \backslash B'$ no player other than Player j has influence over the payoff of Player j. In a quasi-ergodic game there is no problem with a player evaluating its actions locally, as long as the the payoffs in $B \backslash B'$ are not problematic as functions of Player j's actions. In a semi-ergodic game, a player can concentrate entirely on the finite subset B'. Our Bayesian game described below is semi-ergodic.

The difference between the two approaches, their different types of strategies and evaluations, gives an added depth to optimisation and stability. In [5] and [6] we defined a Harsanyi ϵ-equilibrium for a positive ϵ: all players in a Harsanyi ϵ-equilibrium choose measurable strategies with respect to their sigma algebras as defined above and there is no measurable deviation by some player to another measurable strategy resulting in an expected gain of more than ϵ in global evaluation. But there is another type of equilibria, the Bayesian. A Bayesian ϵ-equilibrium is a way for each player to play that is ϵ-optimal for each set in its partition with respect to its local probability distribution on that set. Because the Bayesian equilibrium concept does not require measurable strategies (only that strategies are constant for a player on each set in its partition), there can be Bayesian equilibria where there are no Harsanyi equilibria. This is true for the example in [5], which is also an ergodic game.

There is an added complication to the relation between Harsanyi and Bayesian equilibria when moving to approximate equilibria. Hellman [3] showed that there is a two person ergodic game without a Borel-measurable Bayesian ϵ-equilibria for sufficiently small positive ϵ. For a positive ϵ, a Harsanyi ϵ-equilibrium can be much easier to find than a Harsanyi equilibrium. If the Bayesian game is defined with an amenable structure (for example through the actions of an amenable group or semi-group) there will be a Harsanyi ϵ-equilibrium for every $\epsilon > 0$ even though there may be no Harsanyi equilibrium [4] (as happens with Hellman's example). This is because a Harsanyi ϵ-equilibrium could employ a very small set where the deviation from local ϵ-equilibrium is significant, for example of measure less than $\frac{\epsilon}{B}$ where the deviation in payoff optimality can be no more than B. By performing this deviation the measurable behaviour elsewhere of a Harsanyi ϵ-equilibrium could be supported. The existence of such a small set and its role in supporting a Harsanyi ϵ-equilibrium is related to Folner's condition for amenability.

When moving to finitely additive measures, there are structures to a Bayesian game that enable paradoxical decompositions. To understand this, we take the Bayesian

approach and an ergodic game where the information sets of the players are defined as the orbits of finite groups that generate a non-amenable group G and whose elements are measure preserving. There are two intermediate levels between the whole probability space and individual points in that space. One level is the beliefs of the players as defined by the partitions $\mathcal{P}_j$ and the probability distributions on each partition member. As all sets in all $\mathcal{P}_j$ are finite, this level is very close to the individual points. The other and higher level (involving larger sets) is the collection of subsets that the players know in common, the join partition of the $\mathcal{P}_j$, the largest partition smaller than each of the $\mathcal{P}_j$. The subset that the players know in common, the join partition, is the orbit of G. There may be no probability distribution supported on this set, not even a finitely additive one, that is G-invariant. With the Bayesian approach, the game is played out on these orbits of G. The partitions forming each player's knowledge may be countably generated (the result of countably many refinements of finite partitions of the space) while the join partition may fail to be countably generated. This could frustrate any attempt to create Hansanyi equilibria from Bayesian equilibria, including the broader context of finitely additive measures. Herein lies the special contribution of Bayesian games to measure theoretic paradoxes. Nothing pathological about the information and payoff structures of the individual players (from the finite group actions) is necessary for the game to be paradoxical with respect to equilibria and finitely additive measures.

A Bayesian Game

The following two player Bayesian game has equilibria, yet fails to have measurable ϵ-equilibria for sufficiently small $\epsilon > 0$, where we mean by measurable with respect to any finitely additive extension of the Borel measure that is invariant with respect to the semi-group used to define the information structure of the game.

The most important connection between the information structure of a game and colouring rules is stated above, that if x, y belong to the same information set of a player, then that player must behave identically at x and y. It is this transmission of behaviour over an overlapping system of partitions defined by ergodic operators that connects a colouring rule to equilibrium behaviour.

We use the same space $X' = X \times \{a, b, c\}$ as above. We define two overlapping partitions of $\Omega = X'$ corresponding to two players I and II.

The information sets of Player I are the sets of the form $\{(x, a)\} \cup (T_1^{-1}(x) \times \{b\}) \cup (T_2^{-1}(x) \times \{c\})$. The information sets of Player II are the sets of size three of the form $\{x\} \times \{a, b, c\}$. Player I considers each of these three sets equally likely. Likewise Player II consider each of the three points equally likely.

The matrices B_0 and B_1 and the A_x remain the same, with the same entries $b_{i,j,k}$ of the B_0 or B_1, which B_0 or B_1 determined by x^e, also as above with the F colouring rule.

The payoffs to Player II take place at the points (x, b) and (x, c), with a separate analysis for these two points. Player II has nine actions to choose from the set of (j, k) corresponding to $j = 1, 2, 3$ and $k = 1, 2, 3$. The payoff for Player II is the sum of the payoffs resulting from (x, b) and (x, c); one applies the matrix A_y for $y = T_1 T_2 T_1 x$ to the behaviour of Player I at (x, b) and the matrix A_z for $z = T_2 T_1 T_2 x$

to the behaviour of Player I at (x, c). The maximising of the sum for Player II is accomplished independently. If $p \in \Delta(\{1, 2, 3\})$ is chosen by Player I at (x, b) and $q \in \Delta(\{1, 2, 3\})$ is chosen by Player I at (x, c) (usually different because they come from different information sets of Player I), and Player II chooses the distribution $t = (t_{j,k}) \in \Delta(\{1, 2, 3\} \times \{1, 2, 3\})$ then the payoff for Player II is $r A_y p^t + s A_z q^t$, where y, z are defined as above, $r_j = \sum_k t_{j,k}$ is Player II's marginal distribution on the first coordinate j and $s_k = \sum_j t_{j,k}$ is Player II's marginal distribution on the second coordinate k.

The payoff to Player I takes place entirely at the point labeled a. The payoff for Player I at the ith row is $\sum_{j=1}^{3} \sum_{k=1}^{3} t_{j,k} b_{i,j,k}$, where as above $t_{j,k}$ is the probability that Player II chooses the combination of j with k. In other words, only the j coordinates are relevant at (x, b) and only the k coordinates are relevant at (x, c). The $3 \times 3 \times 3$ matrix is replaced by a 3×9 matrix.

Notice how the causation of players' actions follows through the space. However Player I behaves at a point (x, a) this behaviour gets translated identically to $T_1^{-1} x \times \{b\}$ and to $T_2^{-1} x \times \{c\}$. Player II follows suit at $T_1^{-1} x \times \{b\}$ and at $T_2^{-1} x \times \{c\}$ with an attempt to copy Player I's action. However these points belong in uncountably many different information sets of Player II. When these various actions of Player II happen at a common point (y, a) (translated from both (y, c) and (y, b)), Player I responds accordingly and transmits this response further to $T_1^{-1}(y) \times \{b\}$ and $T_2^{-1}(y) \times \{c\}$. The structure of the informations sets mirrors precisely the structure of descendants used to define the colouring rule F (of Theorem 1 and Theorem 2).

Any measure μ conforming to the local beliefs of the players (making them regular conditional probabilities) must be G' invariant. For all six elements of S_3 this follows from the beliefs of Player II. Given any μ measurable set A, $\mu(A \times \{a\}) = \mu(T_1^{-1}(A \times \{b\}))$ follows from the beliefs of Player I followed by $\mu(A \times \{b\}) = \mu(A \times \{a\})$ from the beliefs of Player II. Together with S_3 measure preserving we get that T_1 is measure preserving. A similar argument shows that T_2 must be measure preserving.

Corollary 2 *For small enough positive ϵ the above Bayesian game does not have an ϵ-equilibrium that is measurable with respect to any finitely additive measure invariant with respect to the semi-groups G.*

Proof The colouring rule F (of Theorem 1 and Theorem 2) is already formulated as a three player Bayesian game, once we identity the three players to the three types of points (x, a), (x, b) (x, c) and their evaluations. With these three players, the satisfaction of the rule F is equivalent to an equilibrium and the ϵ-stability property of a colouring is equivalent to an ϵ-equilibrium, so we have already proved the claim for a three player Bayesian game. The difference is that we have a two player game with the evaluation by Player II on two separate points replacing the two separate evaluations. As evaluating and maximising for Player II occurs independently at these two points, maximisation occurs for both points together if and only if it occurs for each one separately, and ϵ-optimality for Player II is equivalent to ϵ-optimality for the collection of points labelled (x, b) and (x, c) in a colouring. Corollary 1 concerns only the colouring of the points (x, a), once any significant deviation from pure colouring of the points (x, b) and (x, c) are marginalised in probability. Therefore that this game

does not have such measurable γ-equilibria for sufficiently small positive γ follows also from Corollary 1.   □

5 Conclusion

We conjecture that paradoxical probabilistic colouring rules exist when defined with group action. The difficulty seems to lie with the analysis of a stochastic process that isn't reducible to a combinatorial argument via purity.

Does every colouring satisfying our colouring rule F, or every colouring satisfying a paradoxical colouring rule, imply the existence of a measurably G-paradoxical decomposition [7] using sets of the sigma algebra generated by the colour classes, the Borel sets, and the action of the semi-group generated by the measure preserving transformations defining the descendants?

Our colouring rule F concerns finitely additive invariant measures that extend the Borel measure. Does there exist a paradoxical probabilistic colouring rule such that every colouring that satisfies it is not measurable with respect to any finitely additive invariant probability measure. The existence of such a rule in the case of group actions would lead to weak paradoxical decompositions (see [7, Thm. 1])

There is a problem with applying our above colouring rule F to the group action context. We could revise the colouring rule to one on $X = \{0, 1\}^G$ with the group $G = C_2 * C_3$, the context of the Hausdorff paradox. The problem is that the colouring of each point could be so cleverly balanced as to allow for the non-purity of colours throughout the countable orbits of G. Putting these various orbits together may still result in a failure of finitely additive measurability, but the argument for this, if true, is opaque. Surrounding a three-cycle with points coloured purely does not stop the colouring of the three-cycle with non-pure colours, as by Brouwer's Fixed Point Theorem there would always be a colouring, not necessarily pure, satisfying the rule inside the three cycle. One would have to argue that such a colouring inside the cycle must influence the colouring of the three points surrounding that cycle, and as a consequence other cycles, such that completion of the colouring throughout the orbit would not be possible in a measurable way. Such an argument is plausible, however seems very difficult.

The above use of the relation $T_1^3 T_2 = T_2 T_1^3$ was introduced due to the trouble caused by the Brouwer Fixed Point Theorem. Assume there is a probabilistic colouring rule where the colours belong to an $r - 1$ dimensional simplex of r extremal colours, the colour is determined by independent elements (no relations) $g_1, \ldots, g_k$ of a semi-group G of measure preserving transformations, and between the identity and any g_i there are only finitely many elements meaning that for every g_i the set $H_{g_i} = \{h \mid g_i = hg^* \text{ or } g_i = g^*h \text{ for some } g^* \in G\}$ is finite. We define H_{g_i} so that it always contains the identity. Assume further that the colouring rule forces the purity (and uniqueness) of a colour function almost everywhere and the rule breaks down into finitely many parts defined by finitely many clopen sets $A_1, \ldots, A_l$ that partition X. Let H be the union of the H_{g_i} with $|H| = n$. For any $y \in X$ consider all the possible joint distributions on the colours of hy for all the $h \in H$. If there are relations between the g_i, some joint distribution starting at some $y = g_i x$ could conflict with another starting at some $y' = g_j x$, and therefore this approach could lead nowhere (as happens with our main

example). Let H' be the minimal members of H, meaning the non-identity members h such that $h \neq h_1 h_2$ for any two other non-identity members of H. Given no relation between the g_i and given any start $x \in X$, from the joint distributions on the colours of hx for all the $h \in H'$ a joint distribution on the colours of hx for all the $h \in H$ would be determined continuously by the colouring rule and the membership of x in the various A_j (according to their probabilities). The Euclidean space on which we then apply Brouwer's fixed point theorem would have dimension $l \cdot (r^n - 1)$ where r is the number of extremal colours. Though that dimension could be very large, it is still finite. A fixed point of joint distributions, each conditioned on membership in the $A_1, \ldots, A_l$ would result. We would be robbed of our best tool to demonstrate that a colouring rule is paradoxical. Is there a way around the Brouwer fixed point theorem that allows for probabilistic paradoxical colouring rules while maintaining free generation by non-invertible generators?

Another approach is to single out one non-invertible semi-group element T and interpret T^{-1} as the passage of one unit of time. The idea is that with every passage of a unit of time there are uncountable variations for continuation, with some conditional probability distribution governing these variations. In this context one could study how colouring develops over time. Is there a paradoxical colouring rule such that with high probability, with respect non-measurable random starts at colourings that don't satisfy the rule, the process will move toward paradoxical colourings? Non-amenability means that there is a large boundary to any rule obedient area that has the potential to destroy rule compliance quickly. On the one hand, local obedience to the rule could be conformistic and self perpetuating. With our above example of a paradoxical colouring rule, we show that if there is global satisfaction of the rule, it cannot be measurable. We don't provide an understanding of how that paradoxical structure could come into existence.

We conjecture that there are two-player ergodic Bayesian games without finitely additive approximate equilibria (note that our example is semi-ergodic). The discovery of such a Bayesian game would answer all the open problems in the conclusion of [6].

In the context of group actions we asked in [7] about the following: if for a finite sequence of sets $A_1, A_2, \ldots, A_n$ there is no G-invariant measure for which all the sets are measurable, does this imply that one of these sets is *absolutely* non-measurable, meaning that there is no finitely additive G-invariant measure such that this set is measurable? This is a very relevant question to game theory. We would like to know whether the non-existence of a joint measure for the behaviour of all the players in equilibrium implies that for at least one player its behaviour cannot be measurable with respect to any private measure (not shared with the other players) that is consistent with the information structure of the game.

One could raise an objection to our above Bayesian game and our conclusion of no G'-invariant equilibria; why does the finitely additive measure need to be G'-invariant? Why does the measure need to respect the local beliefs of the players? Why should those beliefs be regular conditional probabilities? Indeed the question can be asked in more generality against the colouring rule F on which the game is based, but it can be addressed most efficiently in the game context. Indeed Player I receives a non-zero payoff at only one point in each of its information sets and Player II evaluates actions independently at two different points in each of its corresponding information

sets. The observation supporting the objection is that the local optimisation process remains intact when switching to a finitely additive measure that makes all subsets measurable. The problem with this perspective is that the resulting global expected payoff for a player from such a measure would bear no relation to an integration over the space of the expected payoffs as perceived locally by the players. In the context of game theory, the paradox of our Bayesian game is that one can have equilibria or one can have payoffs as player expectations of measurable functions, but not both simultaneously.

There is another answer to the above objection, observed from a simple change to the payoffs of the game, a change that makes the game quasi-ergodic. Let M be a very large positive number. For the (x, a) point the payoff for Player I from the action corresponding to the colour c_2 could be increased for all combinations of columns by M and likewise the action corresponding to c_2 decreased by M uniformly at the set $T_1^{-1}x \times \{b\}$. At the same point, the payoff for Player I for the action corresponding to c_3 could be decreased uniformly at (x, a) by M and likewise the action corresponding to c_3 increased by M uniformly at the set $T_2^{-1}x \times \{c\}$. Lastly so increase the payoff for Player I uniformly by M for the action corresponding to c_1 at $T_1^{-1}x \times \{b\}$ and decrease it uniformly by M at $T_2^{-1}x \times \{c\}$. Likewise we do something similar to the actions of Player II; at (x, a), (x, b) and (x, c) with the result that nothing is changed strategically as long as the players have the same local probability distributions on their information sets, but if a measure is not G' invariant then the strategic choices of the game are changed. In what way can a measure μ be the basis of an ϵ-equilibria when the local beliefs of the players are not regular conditional probabilities with respect to μ? This question could be explored with our example in mind.

We could consider the following three examples of colouring rules, all three based on the action of the same group and the same space. Let $G = C_2 * C_3$ be the group freely generated by τ and σ with $\sigma^2 = e$ and $\tau^n = e$ for $n \geq 3$. There are n colours $c_1, \ldots, c_n$, represented modulo n. The space is $X = \{0, 1\}^G$.

Rule 1:

$x^e = 0$:

(1) if $\tau^{-1}x$ is coloured with c_i, $i \neq n$, and σx is not coloured c_1 then x is coloured c_i,
(2) if $\tau^{-1}x$ is coloured with c_i and either $i = n$ or σx is coloured c_1 then x is coloured

   c_{i+1}.

$x^e = 1$:
if $\tau^{-1}x$ is coloured with c_i then x is coloured c_{i+1}.

Rule 2:
If $x^e = 1$ and $\tau^{-1}x$ is coloured with c_i then colour x with c_{i+1},
Given $x^e = 0$ and $\tau^{-1}x$ coloured c_i:

(1) if σx is not coloured c_1 then colour x with c_1,
(2) if σx is coloured c_1 then colour x with c_{i+1}.

Rule 3:
The number n is 3.
Given $\tau^{-1}x$ is coloured c_i and τx is coloured c_j:

(1) if $i = j$ then colour x with c_{i+1},

(2) if $i = j + 1$, $i + 1 \neq 1$ and σx is coloured c_1 then colour x with c_{i+1},

(3) if $i = j + 1$, $i + 1 = 1$ and σx is not coloured c_1 then colour x with c_{i+1},

(4) in all other cases, colour x with c_j.

All three rules are very interesting, but we are not sure how they could lead to probabilistic paradoxical colouring rules or other interesting colouring structures. It is relatively easy to show that Rule 1 is paradoxical for $n \geq 5$ (from a discrepancy for the probability of the colour c_1). The third rule, defined in [2] and proven there to be paradoxical, would need to be revised to a rule dependent on location in order for it to be the basis of a probabilistic colouring rule.

References

1. Blackwell, D., Dubins, L.: Existence and non-existence of proper, regular, conditional distributions. Ann. Probab. **3**(5), 741–752 (1975)

2. Bounds, M.: Paradoxical Colouring Rules for Probability Spaces, M.Sc. thesis. London School of Economics (2020)

3. Hellman, Z.: A game with no Bayesian approximate equilibria. J. Econ. Theory **153**(C), 138–151 (2014)

4. Hellman, Z., Levy, Y.: Equilibria existence in Bayesian games: climbing the countable Borel equivalence relation hierarchy. Math. Oper. Res. **47**(1), 367–383 (2022)

5. Simon, R.S.: Games of incomplete information, ergodic theory, and the measurability of equilibria. Isr. J. Math. **138**(1), 73–92 (2003)

6. Simon, R.S., Tomkowicz, G.: A Bayesian game without ϵ-equilibria. Isr. J. Math. **227**, 215–231 (2018)

7. Simon, R.S., Tomkowicz, G.: Paradoxical decompositions and finitary colouring rules, preprint: arXiv:2211.05126

8. Tomkowicz, G., Wagon, S.: The Banach–Tarski Paradox, 2nd edn. Cambridge University Press, Cambridge (2016)