Optimisation and Paradoxical Decompositions

Tuğkan Batu, Robert Simon, Grzegorz Tomkowicz

June 7, 2021

Abstract

We present an optimisation problem defined using a measure preserving action by a group $G$ on a compact probability space $X$. We show that this problem cannot be solved in any way measurable with respect to a finitely additive measure that extends the original probability measure and such that the group $G$ remains measure preserving. The proof involves a computer-performed analysis of a stochastic process. On the other hand, there are some non-measurable solutions that induce some paradoxical decompositions of the space $X$. Additionally, we show that, for some sufficiently small positive $\epsilon$, $\epsilon$-stability also cannot be satisfied in such a measurable way. By $\epsilon$-stability, we mean that the expected gains from local deviations considered independently cannot exceed $\epsilon$.

Tuğkan Batu and Robert Simon
London School of Economics
Department of Mathematics
Houghton Street
London WC2A 2AE
e-mail: \{t.batu, r.s.simon\}@lse.ac.uk

Grzegorz Tomkowicz
Centrum Edukacji $G^2$
ul.Moniuszki 9
41-902 Bytom
Poland
e-mail: gtomko@vp.pl
1 Introduction

A common belief is that measure theoretic paradoxes, like the Banach-Tarski paradox, are not relevant to real events. The reasoning is that such paradoxes require the Axiom of Choice (AC), and therefore the exhibition of paradoxical behaviour would negate the fact that we can remain logically consistent after rejecting AC (Note however that there is a class of theorems called shadows of AC which proofs use AC and exist even after rejecting AC; see [2] for the details.)

There are two parts to showing that a decomposition is paradoxical, a part showing that it cannot be done in a measurable way and a part showing its existence. It is the second part that requires some variation of AC; the first part may not employ it. We tend to forget that there are two parts because usually the construction using AC comes first.

In [5], we considered colouring rules such that the allowed colours for a point are determined by location and the colours of finitely many of the point’s neighbours in a graph. The adjacency relation is defined by measure preserving transformations and the transformations of a point $x$ are called its descendants. A colouring rule is paradoxical if it can be satisfied in some way almost everywhere, but not in any way that is measurable with respect to a finitely additive measure for which the measure preserving transformations remain measure preserving. We established several such paradoxical colouring rules and proved that if there are finitely many colour classes then any colouring of a paradoxical colouring rule has colour classes that jointly with Borel sets define a measurably $G$-paradoxical decomposition of the probability space, by which we mean the existence of two sets of different measures that are $G$-equidecomposable (see [5], Thm. 1).

In the conclusion of [5], 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. 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 evaluation of options, with indifference between two options implying indifference between all of their convex combinations. If the colouring rule were stationary, meaning that the rule does not depend on the position, due to the Kakutani Fixed-Point Theorem, there would be a colouring of the space with one uniform colour, hence there is no probabilistic stationary paradoxical colouring rule. This results from the identity mapping from the colour set $C$ to $k + 1$ copies of $C$ and back down to $C$, where $k$ is the number of descendants.

Another inspiration is the widely held belief in economic theory that though one cannot always accomplish optimisation goals through behaviour that is measurable with respect to a countably additive measure, such as the Lebesgue measure, one can do so always with some finitely additive option. The problem with this perspective is that there may be some 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.

Throughout this paper, by a proper finitely additive extension we mean a
finitely additive measure that extends the Borel measure and is invariant with respect to the measure preserving actions used.

We were inspired optimistically by an example from [5] which gave us some hope that we could extend the results in that paper to the problem of optimization with respect to a continuous cost function with a convex set of options. A brief description of that example follows.

Consider $G = F_2$ and the space $X = \{-1, +1\}^G$ acted on by $G$ in the canonical way (through shifting). Let $g_1, g_2$ be the free generators of $G$. Where the $e$ coordinate of an $x \in X$ lies in $\{-1, +1\}$ determines whether an arrow from $x$ should be directed toward the choice of either $g_1 x$ or $g_2 x$ (if $x^e = +1$) or rather toward the choice of either $g_1^{-1} x$ or $g_2^{-1} x$ (if $x^e = -1$). If a point in the space $X$ has two or more arrows directed to it, it is congested; if not, it is uncongested. The rule is to direct an arrow, if possible, toward a point that is uncongested (if not possible or possible in both directions, the rule allows the arrow to be directed in either direction). We showed that if almost all points follow this rule, the set of congested points is a subset of measure zero. Let the degree of $x$ be the total number of potential inward arrows toward $x$, as determined by the choice of $-1$ or $+1$ for the $e$ coordinate of its four neighbours by the four directions $g_1, g_2, g_1^{-1}, g_2^{-1}$. The argument, that the set of congested points is of measure zero when the rule is satisfied, was combinatorial and stochastic. The structure supporting a congested point requires a continuous repetition of three or four degree points in an infinite chain of vertices, however the average degree of a vertex is two. This implied that the rule was paradoxical, because in at least $\frac{1}{16}$ of the space there was no possibility of any arrow moving inward (and that probability is slightly higher than $\frac{1}{16}$, due to some configurations of two or more points with no possibility of arrows coming in toward that configuration). From every point in the space there was one arrow going outwards, but going inwards on the average there could be at most $\frac{15}{16}$ (assuming a measurable structure).

One could try to repeat the argument with a weighted choice between the two options (meaning some distribution $(p_1, p_2)$ with $0 \leq p_1 \leq 1$ and $p_2 + p_1 = 1$), and devise some definition of what is means to be congested and how congested. With a rule that requires pointing the arrow to the less congested option, one could hope to reproduce a situation where the average weight of received arrows would be strictly less than 1. With the discrete choices there is no option between 1 arrow coming in and 2 arrows coming in. With weighted choices, however, using the critical weight of $1 + \frac{1}{16}$ to defined crowded, there are too many ways to distribute weights so that crowdedness gets evenly distributed throughout typical infinite chains of connected vertices. We kept the same idea that the differences between the degrees of vertices is a useful stochastic process, but we had to look for more sophisticated ways to determine how weights should be distributed.

Our idea was that there should be two kinds of crowdedness, a crowdedness at a point receiving weights, what we call ”passive pain”, and another kind of crowdedness at the location from where weights come, what we call ”active pain”. The idea was that if $x$ could direct some weight to $y$ it is not only the passive crowdedness at $y$, the passive pain, that counts at $x$ but the total weight sent from $x$ to $y$ times the passive pain at $y$. Note that the colouring rule for some $x$ cannot be
determined in any way by the colour of \( x \), however we incorporate into the colour of \( y \) a variable that reflects the colour of \( x \). Rather than a purely combinatorial argument, we wanted to show that pain, both active and passive, almost everywhere had to keep on increasing along an infinite chain. As there is a finite upper limit to the level of both kinds of pain, that would mean that this crowdedness or pain could be sustained only in a set of measure zero.

We discovered that two choices for distributing weights was not enough. Though the group generated freely by two independent elements has arbitrarily many independent choices, to keep the structure simple we equated generators with choices. With \( n \) generators, as before, \(-1\) for the \( e \) coordinate of \( x \) means that weight can be sent only in the negative directions (to \( g_1^{-1}x, \ldots, g_n^{-1}x \)) and \(+1\) means that weight can be sent only in the positive directions (to \( g_1^{+1}x, \ldots, g_n^{+1}x \)). Define the degree of a point \( y \) in \( X \) to be the number of directions from which weight can be sent toward \( y \). With \( n \) generators, the degree ranges from \( 0 \) to \( 2n \) with an average of \( n \).

The distribution of degrees is determined by the binomial expansion.

No matter how many generators were used, there is always a possibility for both passive and active pain to decrease. As a general rule, smaller degrees in a chain means an increase in pain, larger degrees a decrease of pain. The break-even degree is exactly \( n \). Conditioned on having reached a point with some edge, adding one for that connection, the average degree is \( n + \frac{1}{2} \). The product rule for determining active pain means that when the degree is \( n - 1 \) the resulting increase of pain overweighs the resulting decrease of pain when the degree is \( n + 1 \). A closely related analogy is the fact that \( \frac{1}{n - 1} + \frac{1}{n + 1} > 2 \frac{1}{n} \); the effect is stronger for \( \frac{1}{n - k} + \frac{1}{n + k} \) when \( 1 < k < n \). But to exploit this influence sufficiently we needed a variety of possible degrees below the average. We found success at \( n = 5 \), after failing at \( n = 2 \) and \( n = 3 \). At \( n = 4 \), some preliminary work suggested some difficulty in formulating a proof. We suspect that it can be done with \( n = 4 \), but not as nicely as with \( n = 5 \).

In the next section we describe the colouring rule. In the third section we show that this colouring rule is paradoxical, given a certain stochastic structure. In the fourth section, we present a computer analysis confirming that stochastic structure. In the fifth section we apply the colouring rule to a problem of optimisation and show that solutions which are \( \epsilon \)-stable for some positive \( \epsilon \) cannot be measurable with respect to any proper finitely additive extension. In conclusion we consider related problems.

## 2 A Probabilistic Colouring Rule

Let \( G \) be the group freely generated by \( T_1, T_2, T_3, T_4, T_5 \), and let \( X = \{-1, 1\}^G \). 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 group action on \( X \), namely \( g(x)^h = x^{gh} \) for every \( g, h \in G \). We use the canonical product topology and probability measure, that giving \( 2^{-n} \) for every cylinder determined by any choice of \(-1\) or \(+1\) for any \( n \) distinct group elements. With this Borel probability measure the group \( G \) is measure preserving.

**Definition 1.** The graph of \( X \) is the subgraph of the orbit graph of the action of \( G \).
on $X$ induced by the edge subset $\{(x, T_i x) \mid x^e = +1\} \cup \{(x, T_i^{-1} x) \mid x^e = -1\}$. The subset of $X$ where $G$ does not act freely has measure zero. Without loss of generality, we will be interested only in those orbits of $G$ and connected components of the graph of $X$ where $G$ acts freely. We orient the graph of $X$ by placing arrows from $x$ to all five of the $T_i(x)$ if $x^e = +1$ and arrows from $x$ pointed to all five of the $T_i^{-1}(x)$ if $x^e = -1$. We define $S(y) := \{x \mid y = T_i x^e(x)\}$ and define $|S(y)|$ to be the degree of $y$ (the number of neighbours in the graph with arrows pointed to $y$).

The graph of $X$ involves two independent structures of arrows. Each point $x$ has a passive and active role, an active role in one structure and a passive role in the other. The active and passive roles alternate. We will be interested in that alternation, moving from a point in its passive role to its neighbours in their active roles, and from a point in its active role to its neighbours in their passive roles. Every point has an active role, namely a connection to five different points in their passive roles. The degree of a point concerns its passive role. Not every point has a passive role; some are of degree zero. The points of degree zero play indirectly a passive role. Not every point has an active role, namely a connection to five different points in their active roles. The active and passive roles alternate. We will be interested in that alternation, moving from a point in its passive role to its neighbours in their active roles, and from a point in its active role to its neighbours in their passive roles. Every point has an active role, namely a connection to five different points in their passive roles. The degree of a point concerns its passive role. Not every point has a passive role; some are of degree zero. The points of degree zero play indirectly a passive role.

Every point $x \in X$ has a colour in $\Delta(\{1, 2, 3, 4, 5\}) \times [0, 1] \times z \in S(x) [0, 1]$, where the dimension of the last part of the colouring is equal to the degree of $x$. We let the first part, $\Delta(\{1, 2, 3, 4, 5\})$, a four dimensional simplex, be called the active part of the colour and the $[0, 1] \times z \in S(x) [0, 1]$ part the passive part of the colour.

What is the colouring rule, which we call $Q$?

Usually the word "cost" is used to describe a function that should be minimised. With this example, we prefer the word "pain", because it represents a situation that should be avoided. In every $i \in \{1, 2, 3, 4, 5\}$ of the five directions we define the active pain for $x$ to be $w \cdot r_x$ where $w$ is the first coordinate of the passive colour of $y = T_i x^e(x)$ and $r_x$ is the coordinate corresponding to $y \in S(x)$ in the passive colour of $y = T_i x^e(x)$. The rule for the active colour of $x$ is to choose those directions where the active pain is minimised. If more than one are minimal, then any convex combination of the minimal directions is allowed. The quantity of the active colour of $x$ given to the $i$ coordinate (in the direction of $T_i$ or $T_i^{-1}$) is called the weight given in the direction $i$ or toward $y = T_i x^e(x)$.

The first part of the passive colouring is called the passive pain. The rule for the first part of the passive colouring is as follows. First an $r$ is chosen for the whole space such that $0 < r < 2^{-10}$. Whenever the sum of the active colours in $S(y)$ moving toward $y$ is less than $1 + r$, then $w = 0$ is required by the rule $Q$ for the first passive coordinate. If that sum is more than $1 + r$ then $w = 1$ is required by the rule $Q$. And if the sum is exactly $1 + r$ then any value in $[0, 1]$ is acceptable for $w$.

The rule for the $x \in S(y)$ coordinate of the passive colour is very simple, it is the copy of the $i$ coordinate of the active part of the point $x$ pointed toward $y$ such that $y = T_i x^e(x)$.

It is now clear from the colouring rule $Q$, why $T_i$ should remain measure preserving with any finitely additive extension. A critical aspect of the colouring rule uses that from any $y$ all the $x$ such that $x \to y$ are treated equally, e.g. their weights are summed without prejudice.
First, it is easy to show, with AC (for uncountable families of sets since there are uncountably many group orbits), that there is some non-measurable solution to the rule \( Q \) valid almost everywhere. By AC we can choose in each orbit where \( G \) acts freely a special point \( x \) to correspond to \( e \in G \). Classify each point \( y \) in the orbit containing \( x \) according to the length of the word in \( G \) needed to move from \( x \) to \( y \). Choose a direction from \( y \) that involves a word of one length greater and allowed by the coordinate \( y^e \). As there is only one possibility for a direction from \( y \) corresponding to a word of one length less (and no such possibility if \( y = x \)), there will always be an option to satisfy this requirement. Because one always chooses an arrow from a point with a shorter word to one with a longer word, it is not possible for two arrows to be pointed toward the same point. The end result will be a colouring satisfying the rule \( Q \) where there is no pain, passive or active. As the solution is deterministic, e.g. all weight goes toward one direction, by [5] from this solution one can construct a measurably \( G \)-paradoxical decomposition.

3 Paradoxical Colouring

The goal of this section is to complete the proof of the following theorem:

**Theorem 1.** The colouring rule \( Q \) is paradoxical.

We have already shown that there is some way to satisfy the rule. To complete the proof of Theorem 1 we will show that satisfaction of the colouring rule implies that the passive and active pain is 0 almost everywhere (with respect to the Borel measure). To show that \( Q \) is paradoxical, it suffices to show that the set of points where the passive pain is equal to 1 is contained in a set of measure strictly less than \( 2^{-10} + 2^{-10} \). That would be enough to show that the average weight moving toward all points is strictly less than 1.

**Definition 2.** Given that \( x \rightarrow y \), meaning that \( y = T_{x^e}x \) if \( x^e = +1 \) or \( y = T_i^{-1}x \) if \( x^e = -1 \), the chain generated by \( x \rightarrow y \) are all the points \( z \) in the graph of \( X \) that can be reached from \( x \) without going through \( y \) and involve alternating arrows, meaning that if \( z \) is an odd distance from \( x \) then the passive role \( z \leftarrow \) connects \( z \) to \( x \) and if \( z \) is an even distance from \( x \) then the active role \( z \rightarrow \) connects \( z \) to \( x \). A chain involves an alternating process of using the active and passive colouring functions of the points. If \( x \rightarrow y \) then \( x \) distributes weights to four other points \( z_1, z_2, z_3, z_4 \). In turn, depending on their degrees, there are further directed edges \( x^* \rightarrow z_i \) (or in the rare case that the degree of each \( z_i \) is 1, no further directed edges). One could see a chain as a quarter or less of a \( G \) orbit; the choice of seeing \( y \) as passive or active, and the choice of moving in the \( x \) direction rather than in the direction of the potentially other \( x^* \) with \( x^* \rightarrow y \).

**Definition 3.** A 1st level terminating point of a chain is some vertex \( x^* \) of the chain such that \( x^* \rightarrow z \) is a directed edge of the chain and \( z \) has degree 1, meaning that there is no other \( \hat{x} \) in the chain such that \( \hat{x} \rightarrow z \). That vertex \( z \), and all vertices of degree 1 in the chain, is a terminating point of level 0. If \( i \geq 2 \) is even then an \( i \)th level terminating point is some vertex \( z \) such that \( x^* \rightarrow z \) is a directed edge of
the chain, \( x^* \) is an \( i - 1 \) level terminating point with \( x^* \to z^* \) for some \( i - 2 \) level terminating point \( z^* \). If \( i \geq 3 \) is odd then an \( i \)th level terminating point of a chain is some vertex \( x^* \) of the chain such that \( x^* \to z \) is the last step in the alternating path from \( x \) to \( z \), there is some other \( \hat{x} \neq x^* \) such that \( \hat{x} \) is a terminating point of level \( i - 2 \) with \( \hat{x} \to z \) and all the other \( \bar{x} \) such that \( \bar{x} \to z \) with \( \bar{x} \neq x^* \) and \( \bar{x} \neq \hat{x} \) are terminating points of even level equal to or less than \( i - 2 \). A terminating point is a vertex that is a terminating point of some level. A chain generated by \( x \to y \) is terminating if \( x \) is a terminating point, and its terminating level is the terminating level of \( x \). If the chain generated by \( x \to y \) is not terminating, then we say that the chain and the edge \( x \to y \) is non-terminating. Notice that this does not imply that \( y \) is a terminating point, as there could be at least two edges \( x^* \to y \) with \( x^* \neq x \) such that \( x^* \to y \) is not a terminating chain. The non-terminating part of a non-terminating chain is the non-terminating chain with its terminating points removed.

Remark: A subchain of a terminating chain may not be a terminating chain. A subchain of a non-terminating chain may be a terminating chain. We could have a terminating chain generated by \( x \to y \) with \( x \to z_1 \), \( x \) terminating of level 1, \( z_1 \) terminating of level 0, and \( x \to z_2 \) with \( x^* \to z_2 \) generating a non-terminating chain for some \( x^* \neq x \). Likewise \( x \to y \) could be non-terminating, \( x \to z \) with \( z \) of degree three or more, with \( x^* \to z \) generating a terminating chain for some \( x^* \neq x \).

**Lemma 1.** If \( x \to y \) generates a terminating chain, then in any colouring of \( X \) that satisfies the colouring rule \( Q \), the active pain level at \( x \) is zero, meaning that if the passive pain level at \( y \) is positive then no weight is given at \( x \) toward \( y \).

**Proof.** We prove the lemma by induction on the the level of the terminating point; we claim that any terminating point of odd level experiences no passive pain and any terminating point of even level experiences no active pain. Suppose that \( x^* \to y^* \) is part of the chain and \( y^* \) is terminating of level 0. As \( y^* \) is degree one, it is not possible for \( y^* \) to experience passive pain, since the maximal weight sent to \( y^* \) is at most 1. By choosing any positive quantity of weight, including all weight, to \( y^* \), there is no resulting active pain, and since the rule \( Q \) requires that \( x^* \) minimise the active pain, \( x^* \) cannot experience active pain. We notice that such an \( x^* \) is a terminating point of level 1, and so we can carry on with induction. Suppose \( y^* \) is a terminating point of even level \( i \) with \( x^* \) the vertex such that \( x^* \to y^* \) and \( x^* \) and the path from \( y \) to \( y^* \) passes through \( x^* \). If \( y^* \) was experiencing any passive pain, all the points \( \hat{x} \to y^* \) such that \( \hat{x} \neq x^* \) would give zero weight to \( y^* \), since by induction they all experience no active pain. And with only one vertex \( x^* \) possibly giving weight to \( y^* \), it is impossible for \( y^* \) to experience any passive pain. But then \( x^* \) does not experience any active pain, because it could put all weight toward \( y^* \).

We describe a structure essential to our following stochastic arguments. Instead of looking at some \( x \) according to its topological location in \( X \), we think of \( x \) as a member of a chain. Given that \( x \) sends the weight \( p > 0 \) to \( y \) with \( t > 0 \) the passive pain level at \( y \), we consider what the passive pain levels \( t_i \) must be at the four \( z_i \) with \( x \to z_i \) for all \( i = 1, 2, 3, 4 \) so that the active pain levels for each of the five choices are equal (through satisfying the rule \( Q \)). If \( p_i \) is the weight sent from \( x \) to \( z_i \), we
must have the equations \( p_i t_i = pt \) for each \( i = 1, 2, 3, 4 \). With the equations \( t_i = \frac{n_i}{z_i} \), we also consider the degrees of the \( z_i \) and how these quantities continue in further stages in the chain generated by \( x \rightarrow y \). We want to show that almost everywhere, given \( t > 0 \), the pain values in further stages is unbounded. Since these values cannot exceed 1, we have shown that positive passive pain happens only in a subset of measure zero. It is a kind of reverse engineering, determining what pain values must exist as implied by the rule \( Q \). In this analysis we do not focus on directional choices determined by the \( e \) coordinate; we look instead on the degrees of the vertices of odd distance to \( x \) (even distance from \( y \)). We use that the probability of degree \( k \) is \( \binom{9}{k-1} 2^{-9} \). It does not follow the formula \( \binom{10}{k} 2^{-10} \) because we condition on the existence of a particular edge \( x \rightarrow y \).

A first step toward the main argument is to eliminate all terminating chains from the analysis and look at only non-terminating chains and their non-terminating parts. The following lemma does this.

**Lemma 2.** The probability that \( x \rightarrow y \) does not generate a terminating chain is approximately \( \hat{q} = .991603 \).

**Proof.** Because of the homogeneous structure to the space, there is a recursive formula for the value of \( \hat{q} \). Given that \( x \rightarrow z \) and \( z \neq y \) and \( x^* \rightarrow z \) with \( x^* \neq x \), \( \hat{q} \) is also the probability that \( x^* \) is not a terminating point. The probability that \( z \) is a terminating point is \((1 - \frac{\hat{q}}{2})^9\), hence the probability of it not being a terminating point is \(1 - (1 - \frac{\hat{q}}{2})^9\). For \( x \) to not be a terminating point each of the four such \( z \) must fail to be terminating points. Therefore \( \hat{q} \) is the root of the polynomial \( q = (1 - (1 - \frac{\hat{q}}{2})^9)^4 \). Applying Wolfram Alpha, the largest root of this polynomial strictly less than 1 is \( \hat{q} = .991603 \).

In what follows, we will assume that all terminating points are removed so that the probability distribution on the degrees in a chain follow the binomial expansion applied to \( \frac{q}{2} \) and \( 1 - \frac{q}{2} \) (instead of \( \frac{1}{2} \) and \( \frac{1}{2} \)) and then conditioned to probability \( \hat{q} \).

**Definition 4.** Given a non-terminating chain generated by \( x \rightarrow y \), any \( p \in (0, 1) \), and any colouring \( c \) of that chain satisfying the rule \( Q \) with \( p \) the weight of \( x \) given to \( y \) and \( 1 \) the passive pain at \( y \), define \( t(x \rightarrow y, c, p) \) to be the supremum of the active pain in the colouring \( c \). Because we are concerned with the ratio to the quantity 1, e.g., the passive pain at \( y \) could be some very small positive \( v \), we allow for values above 1, although strictly speaking there can never be pain, passive or active, above the level of 1. By definition, \( t(x \rightarrow y, c, p) \geq p \). Define \( u(x \rightarrow y, p) \) to be the infimum of \( t(x \rightarrow y, c, p) \) over all the colourings \( c \) satisfying the rule \( Q \). It is straightforward that \( u(x \rightarrow y, p) \) is increasing in \( p \). Due to the finitely many possibilities for vertices on each stage of the process, \( u(x \rightarrow y, p) \) is a Borel measurable function of \( x \rightarrow y \) and \( p \). Lastly for every \( p \) define \( u(p) \) to be the greatest lower bound for the set \( \{ w \mid u(x \rightarrow y, p) \geq w \text{ almost everywhere} \} \). Because it is non-decreasing, \( u(p) \) is also Borel measurable.

Our goal is to prove the following proposition.

**Proposition 1.** The function \( u(p) \) is infinite for all \( p > 0 \).
Proposition \( \Box \) implies Theorem \( \Box \). This follows directly from the definition of \( u(p) \). As both passive and active pain cannot exceed 1, Proposition \( \Box \) implies that the places where the passive pain is 1 is a set of measure zero. To prove Theorem \( \Box \) it suffices to show that \( u(p) \) is infinite for all \( p > \frac{1}{100} \), as any point with passive pain 1 must be next to some vertex pointed toward it with weight more than \( \frac{1}{10} \).

Let’s look again at the chain generated by \( x \to y \), where \( x \to z_i \) for \( i = 1, 2, 3, 4 \) and \( x^* \) is another vertex where \( x^* \to z_i \neq y \) for some \( i \). Assume that \( p \) is the weight given by \( x \) toward \( y \). The function \( u(\to y, p) \) is defined by a minimisation of the \( u(x^* \to z_i, q) \cdot v_i \) over all the choices for weights \( q \) from \( x^* \neq x \) to the various \( z_i \) and by the weights sent to the \( z_i \) from \( x \) and the corresponding induced passive pain levels \( v_i \) for the \( z_i \), with equality to \( u(\to y, p) \) given that the maximum is not already obtained with \( p \) times the passive pain level at \( y \), which we normalise to 1 to define \( u(x \to y, p) \). This means that the \( u(x \to y, p) \) times the induced passive pain, when the values are finite, form a super-martingale through the minimisation process (potentially increasing future values through the minimisation happening at the start). Instead of doing this on each stage, we will minimise instead using the function \( u(q) \) instead of \( u(x^* \to z_i, q) \) – given that \( q \) is the weight from a vertex \( x^* \) toward \( z_i \) with \( x^* \neq x \), we make the assumption that the maximal active pain in the chain generated by \( x^* \to z_i \) is \( u(q) \) times the passive pain of \( z_i \). Of course the actual process of pain minimisation could look very different because \( u(x^* \to z_i, q) \) may be much larger than \( u(q) \). But if we minimise with this assumption, we obtain a result which is not higher than the proper result almost everywhere. Our goal is to show that with this assumption, by applying the rule of minimising the \( u(q) \cdot v_i \) level on the next stage, and applying this rule stage after stage, the resulting limit superior of the product values is infinite almost everywhere, and therefore the maximum must be infinite almost everywhere without this assumption.

There are three useful facts about the above assumption that the future maximal active pain level on the next stage is determined by the function \( u \).

First, if one aims to minimise active pain level according to \( u(q) \) times the passive pain level of \( z_i \) on the next stage (where \( q \) is the weight given by some \( x^* \) to \( z_i \)), because \( u \) is a non-decreasing function, it suffices to uses equal weights \( q \) from all the \( x^* \neq x \) such that \( x^* \to z_i \).

The second fact, following from the above equal weights consequence, is that we do not have to consider any \( q \) such that \( q < \frac{1}{100} \). As we already assumed that all the weights on the other side of \( z_i \) were equal, if \( q < \frac{1}{100} \) were these weights it means that the weight from \( x \) to \( z_i \) is at least .91. This means that the probabilities from \( x \) to the other \( z_j \) with \( z_j \neq y \) add up to no more than 0.09. Therefore the passive pain level at these other \( z_j \neq z_i, y \) are at least 10 times that of \( z_i \). Assuming that this is not a terminating chain, there is a \( \pi \) giving weight of at least \( \frac{1}{10} \) to one of these \( z_j \). From the monotonicity of \( u \) the chain \( \pi \to z_j \) receives at least ten times the level of the chain \( x \to z_i \). This means that we could redistribute the weights coming from \( x \), giving more to the other \( z_j \) and less to \( z_i \) with a reduction in the level as determined by the function \( u \). We discovered from our computer calculations that this bound of 0.01 could be replaced by 0.055, but we will continue to use 0.01 in our arguments.

The third fact from the above assumption is that the products of \( u(q) \) times
the passive pain level \(v_i\) of its corresponding \(z_i\) are minimising through equalising (which also gives a unique result). If that product is largest for the branch after \(z_i\) and smallest for the branch after \(z_j\) one can move some weight given by \(x\) away from \(z_j\) and toward \(z_i\). That would reduce the passive pain at \(z_i\), increase it at \(z_j\).

We do not know what is the function \(u\) explicitly, so we go one step further; instead of minimising according to the monotonically non-decreasing function \(u\), we introduce some new function \(w_0\), also monotonically non-decreasing, to be minimised on the next stage instead of the \(u\). Though we want it to be never larger than \(u\), what really matters is that it is no larger than \(u\) or \(p\) on all \(p\) greater than \(\frac{1}{100}\). As \(u(p) \geq p\), it suffices to say that \(w_0(p) \leq p\) for all \(p\). We do define it normalised according to a passive pain level of 1 given to \(y\), so that if that passive pain is \(v\) then the \(w_0(p)\) gets multiplied by \(v\) if \(p\) is the weight given to \(y\) by \(x\). When minimising with respect to \(w_0\) on the next stage, we keep the three consequences from the above assumption, as they were only consequences of the fact that \(u\) is defined for all chains in the same way and that it is monotone. In particular, on the next stage we do not have to consider values for \(q\) lower than \(\frac{1}{100}\). Because of that application of \(w_0\), in what follows we assume it is not defined for \(p\) less than \(\frac{1}{100}\). We will show that the process of stage for stage minimisation of the product of the \(w_0(p)\) with the induced passive pain levels will result in an infinite limit superior almost everywhere.

Let \(x \rightarrow y\) define a non-terminating chain and let \(B\) be the vertices of this chain of distance no more than 2 from \(x\) in the graph. Let \(B'\) be the vertices of exactly distance 2 from \(x\) and let \(z_1, z_2, z_3, z_4\) the vertices other than \(y\) of distance exactly 1 from \(x\). Let \(w_0\) be any non-decreasing positive and bounded real valued function defined on \([\frac{1}{100}, 1]\) such that \(w_0(p) \leq p\) for all \(p\) and there is some \(\frac{1}{2} < \alpha < 1\) with \(w_0(q) = w_0(\alpha)\) for all \(q > \alpha\). For every choice of \(1 \leq j_i \leq 9\) and positive \(s_i\) for \(i = 1, 2, 3, 4\) such that \(p + \sum_{i=1}^{4} s_i = 1\) define \(w_1(p, s_1, s_2, s_3, s_4, j_1, j_2, j_3, j_4)\) to be the maximum over \(i = 1, 2, 3, 4\) of the \(w_0(\frac{1-p}{j_i} s_i)\). The \(s_i\) stand for the weights given by \(x\) to \(z_i\), the \(\frac{p}{s_i}\) stand for passive pain of the \(z_i\) induced by the rule \(Q\) as applied to \(x\), \(y\) and the \(z_i\), and the \(j_i\) stands for the degree of \(z_i\) minus 1. Define \(w_1(p, j_1, j_2, j_3, j_4)\) the minimum of the \(w_1(p, s_1, s_2, s_3, s_4, j_1, j_2, j_3, j_4)\), minimised over the choices for \(s_1, s_2, s_3, s_4\). Notice that the use of a small positive quantity \(r\) is not involved. It would only increase the values, as it would increase the weights on the other side of the \(z_i\). Keeping with \(r = 0\), we show that the \(u\) function is a.e. infinite. Later when we consider \(\epsilon\)-stability a positive value for \(r\) will return.

These calculations are not too difficult to carry out, because of monotonicity properties. If more weight is sent from \(x\) to \(z_i\) then in calculating the needed passive pain level at \(z_i\) the reciprocal of the weight is used for that calculation. And if more weight is sent to \(z_i\) then there is less weight to distribute to the other side. The effect from both influences is toward a lowering of the value for that branch of the chain. This means that the task of minimisation is the same as equalisation and the solution is unique. Although the task of solving four non-linear simultaneous equations cannot be done precisely, for a computer the task of finding an approximate solution is an easy one.

As before, let \(\hat{q}\) be the probability that a chain is terminating, which we approximated at \(\hat{q} = .991603\). For each choice of \(1 \leq j_1, j_2, j_3, j_4 \leq 9\), we multiply the
logarithm of \(w_1(p, j_1, j_2, j_3, j_4)\) by the probability

\[
\prod_{i=1}^{4} \left( \frac{9}{j_i} \right) \left( \frac{\hat{q}}{2} \right)^{j_i} \left( 1 - \frac{\hat{q}}{2} \right)^{9-j_i}
\]

and call this \(\tilde{w}_1(p, j_1, j_2, j_3, j_4)\). Function \(w_1(p)\) is defined to be the exponential of \(\frac{4}{q} \sum_{1 \leq j_1, j_2, j_3, j_4} \tilde{w}_1(j_1, j_2, j_3, j_4)\). By the definition of \(\hat{q}\), we have

\[
\hat{q} = \sum_{1 \leq j_1, j_2} \prod_{i=1}^{4} \left( \frac{9}{j_i} \right) \left( \frac{\hat{q}}{2} \right)^{j_i} \left( 1 - \frac{\hat{q}}{2} \right)^{9-j_i},
\]

so we are conditioning on the event that the chain in not terminating.

As we are evaluating a product whose values are always positive, it is the logarithm that matters. For example, if at one stage, we drop uniformly by \(\frac{1}{4}\) and then rise uniformly on the next stage by 2 we have broken even. The idea is to choose a function \(w_0\) such that \(u(p) \geq p \geq w_0(p)\) for all \(p \in \left[ \frac{1}{100}, \alpha \right]\) and show that if minimising \(w_0\) times the induced passive pain stage for stage has an infinite limit superior for all such choices for \(p\) then \(u(p)\) must also be infinite for all such \(p\), as the supremum value is no less than the limit superior value. Any such function \(w_0\) with \(w_0(\alpha) < \infty\) suffices, since we can scale \(w_0\) down so that it is below \(u\). The challenge is to find a \(w_0\) and a positive \(\rho\) such that \(w_1 \geq w_0(1 + \rho)\) for all choices of \(p\).

We have not yet established that the stage for stage minimisation of the \(w_0\) times induced passive pain on a chain generated by \(x \rightarrow y\) is no greater than the values for the same when defining the value \(u(x \rightarrow y, p)\). To do this, we introduce a formalistic approach to those values. We can treat any non-comparable subset of non-terminating \(x^* \rightarrow y^*\) in a non-terminating chain (generated by \(\rightarrow y\)) as a kind of stop function. By non-comparable we mean that there are no two \(x_1 \rightarrow y_1\) and \(x_2 \rightarrow y_2\) in this subset such that \(y_2\) follows \(x_1\) in the path between \(y_1\) and \(x_2\).

At any \(x^* \rightarrow y^*\) in this non-comparable subset the function \(u(x^* \rightarrow y^*, p)\) will be replaced by \(u^*(x^* \rightarrow y^*, p)\) and all calculations of earlier \(u(\hat{x} \rightarrow \hat{y}, p)\) are calculated with these future changes (meaning that the weights will be altered to create a new minimisation).

By changing the future evaluations of some \(x^* \rightarrow y^*\) in \(S\), the \(Q\) rule remains in effect for all the points in the chain generated by \(x \rightarrow y\) that occurred before getting reaching \(x^* \rightarrow y^*\). More precisely the level of passive pain at \(y^*\) remains determined by the \(\hat{x} \rightarrow y^*\) that came before, meaning \(\hat{x}\) is on the path between \(y\) and \(y^*\). Also the \(Q\) rule remains in force at \(y^*\) for the process of copying the weight from \(\hat{x}\) to \(y^*\). With the introduction of the \(x^* \rightarrow y^*\) from \(S\) and its new evaluation, what was minimal with respect to \(Q\) may no longer be minimal and therefore needs to be changed. It doesn’t mean however that the rule \(Q\) ceased to be obeyed.

**Lemma 3.** Let \(S\) be a non-comparable subset of non-terminating \(x^* \rightarrow y^*\) in the non-terminating chain generated by \(x \rightarrow y\). If \(u^*(x^* \rightarrow y^*, p) \leq u(x^* \rightarrow y^*, p)\) for all \(p\), then \(u^*(\hat{x} \rightarrow \hat{y}, p) \leq u(\hat{x} \rightarrow \hat{y})\) for all non-terminating \(\hat{x} \rightarrow \hat{y}\).

**Proof.** By induction it suffices to show this for any \(x^* \rightarrow y^*\) following directly after \(\hat{x} \rightarrow \hat{y}\), meaning that \(y^*\) is a member of the subset \(Z\) of four vertices following \(\hat{x}\) and
not equal to \( \hat{y} \). Let \( p \) be fixed as the weight given by \( \hat{x} \) to \( \hat{y} \). Let the \( q_\text{cr} \) be the weights given to the \( y^* \) by the corresponding \( \overline{P} \), where one of those \( \overline{P} \) is \( x^* \). By decreasing to \( u^*(x^* \rightarrow y^*, p) \) for every \( p \) the corresponding value \( t^*(\hat{x} \rightarrow \hat{y}, p) \) cannot rise, where the \( t^* \) stands for the evaluation at \( \hat{x} \rightarrow \hat{y} \) according to the \( u^*(x^* \rightarrow y^*, p) \) without any change in the weights and without the assumption of minimisation. Whatever minimises to define \( u^*(\hat{x} \rightarrow \hat{y}, p) \) cannot be higher that what existed before, as \( t^*(\hat{x} \rightarrow \hat{y}, p) \) corresponds to some solution, not necessarily a minimal one. \( \blacksquare \)

Next we define the \( w_0 \) process on the non-terminating part of a non-terminating chain generated by \( x \rightarrow y \). Starting at any large even \( N \) of vertices \( x^* \) of distance \( N \) from \( x \), all the \( u(x^* \rightarrow y^*, p) \) are replaced by the function \( w_0 \) and values \( u^*(\hat{x}, \hat{y}, p) \) are created for the other non-terminating \( \hat{x} \rightarrow \hat{y} \). If \( p \) is less than \( \frac{1}{100} \) it does not matter how \( w_0(p) \) is defined as long as the \( w_0 \) remains monotonic, because that probability will be revised later. After making those adjustments we do the same with the vertices of distance \( N - 2 \) from \( x \). Finally we replace \( u^*(x \rightarrow y, p) \) with \( w_0(p) \) for all \( p \) (and if \( p, \frac{1}{100} \) then it does not matter how \( w_0(p) \) is defined in the same way as above). Notice that it does not matter which \( N \) we start with, the results for all stages will be the same. The key to the process is that at each stage the rule for determining a passive pain level is being followed, and we define \( w_0(\hat{x} \rightarrow \hat{y}, p) \) to be the limit superior (over the stages) of the maximal values of the \( w_0(x^* \rightarrow y^*, q^*) \) times the passive pain level at \( y^* \) at each stage induced by the last minimisation giving 1 to the passive pain level for \( y \). In this way \( w_0(x \rightarrow y, p) \) is also defined. Notice that the function \( w_1(p) \) is the expected value for the \( w_0(x^* \rightarrow y^*, q^*) \) times the induced passive pain at \( y^* \) on the next stage, the expectation over all the non-terminating choices.

**Lemma 4.** If \( x \rightarrow y \) is non-terminating and \( u(x \rightarrow y, p) \) is finite, then \( w_0(x \rightarrow y, p) \leq u(x \rightarrow y, p) \).

**Proof.** This follows by Lemma 3 the assumption that \( w_0(x \rightarrow y, p) \leq u(x \rightarrow y, p) \) for all \( p \), and that the supremum is never less than the limit supremum. \( \blacksquare \)

**Proposition 2.** Let \( w_0 \) and \( w_1 \) be functions satisfying the conditions stated above. If there is some \( s > 0 \) such that for every \( p \), \( \log(w_1(p)) > \log(w_0(p)) + s \) then \( u(p) \) must be infinite for all \( p > 0 \).

**Remark:** After a demonstration of such a function \( w_0 \), Proposition 2 proves Proposition 1 (which proves Theorem 1).

**Proof.** We define a Markov chain on the state space \( \mathbb{R} \times [\frac{1}{100}, \alpha] \) according to the \( w_0 \) process. The Markov chain starts at \((0, p)\), where 0 is the logarithm of 1, which we assume to be the passive pain at \( y \) and \( p \) is the weight given to \( y \) by \( x \) (but it does not matter what value for the passive pain we start with). For each choice of degrees \( j_1 + 1 \) for the \( z_i \) and a choice of \( p \), a transition is determined by the process of defining the \( w_1(p, j_1, j_2, j_3, j_4) \), the process of minimising the values on the next stage. The transitions of the Markov chain are determined by equal probability given to the \( z_i \) (it does not matter how the transitions to the vertices are defined since we equalise the active pain results) and corresponding to the various choices.
of the $j_i$, as determined by the binomial expansions using $\hat{q}$ and conditioned by $\hat{q}$ as done above. These weights are determined uniquely by the minimisation process. The next stage is determined by a start at the $x^* \rightarrow z_i$, namely the pairing of the passive pain of the $z_i$ induced by the rule $Q$ applied to $x$ and the $z_i$ with $q_i$ the equal weights toward $z_i$ from those $x^* \neq x$ with $x^* \rightarrow z_i$. If the process requires some $q_i$ on the next stage greater than $\alpha$, simply redefine it downwards to be $\alpha$. As there are only finitely many possibilities for degrees and the process of determining optimality is unambiguous with no weights less than $\frac{1}{100}$, the Markov chain is well defined. This defines a sub-martingale (potentially increasing values in expectation as the stages increase) where the sub-martingale function at $(r, p)$ on the $n$th stage is $\log(r \cdot w_0(p)) - sn$. We can adjust the first coordinate (representing the passive pain) downward on each stage to define a martingale without making any change in the transitions, as the first coordinate plays no rule in defining the transitions.

The Chebyshev martingale inequality states that if $X_1, X_2, \ldots$ is a martingale starting at $X_0$ then for every $\epsilon > 0$ the probability that $|X_n - X_0| > \epsilon$ is no more than the sum of the variances of the $X_i - X_{i-1}$ divided by $\epsilon^2$. As $w_0$ is a function with positive values bounded between $w_0(\frac{1}{100})$ and $w_0(\alpha)$ and all $q$ are between $\frac{1}{100}$ and $\alpha$, the possible variances at each stage have a uniform bound $B$ (determined by the two extremes of all the $z_i$ of degree 2 and all of the $z_i$ of degree 10). As it is a martingale, the cumulative variance of the process to the $n$ stage is the sum of the variances at each stage, which is no more than $nB$. We show that the limit superior over the chain (the function $w_0(x \rightarrow y, p)$) is infinite almost everywhere. In order for a chain $x \rightarrow y$ to fail to have an infinite limit superior value there must be some $N_x$ such that $|X_n(x)| > \frac{\alpha \epsilon}{\epsilon^2}$ for all $n \geq N_x$. Due to countable additivity of the Borel measure, for every $\delta > 0$ there is some universal $N$ such that $N_x \leq N$ for all but a subset of such $x$ (with finite limit superior) of measure $\delta$. Therefore if the subset where the limit superior is not infinite has positive measure, there must be an $\epsilon > 0$ such that for every $n$ the probability that $|X_n| > \frac{\alpha \epsilon}{\epsilon^2}$ is at least $\epsilon$. But this is not true, since the Chebyshev martingale inequality says that this probability is not greater than $\frac{4nB}{n^2 \epsilon^2}$ for every $n$.

Finally, by Lemma 4 the $u(x \rightarrow y, p)$ can finite in at most a Borel set of measure zero, which implies that $u(p)$ is infinite almost everywhere for all positive $p$.

With the Kolmogorov martingale inequality we could have shown that the limit inferior is also infinite a.e., but we did not need it.

In the next section, we show that the rule $Q$ satisfies Proposition 2 using the functions $w_0$ and $w_1$ as determined by a computer analysis.

## 4 The Numerical Calculations of $w_0$ and $w_1$

Where does the function $w_0$ come from and how is the minimal ratio $\frac{w_1(p)}{w_0(p)}$ calculated from the $w_0$? We consider a non-terminating chain generated by $x \rightarrow y$. We assume that the passive pain at some point $y$ is normalised at 1, the probability coming from $x$ to $y$ is $p$, and the non-terminating vertices of even distance from $y$ all have degree 5. This means that, if $p = \frac{1}{5}$, the resulting passive pain of the entire non-terminating part of the chain has the value of 1 and the active
pain \( \frac{1}{2} \). Following from \( x \) to the four other vertices on the other side from \( y \), we assume that the weights are distributed evenly. This means that from \( x \) to the four points \( z_1, z_2, z_3, z_4 \), the weight is \( \frac{1-p}{4} \). The passive pain \( v \) at each of these four places \( z_i \) has to satisfy \( \frac{1-p}{4}v = p \) or \( v = \frac{4p}{1-4p} \). We continue with weights \( \frac{34p}{16} \) in the further directions from each of the \( z_i \). The recursive calculation, a generating function, for the active pain in the limit becomes easier with the substitution \( t = p - \frac{1}{8} \):

\[
g(t) = \frac{1 + 5t}{1 - \frac{5t}{4}} \cdot \frac{1 + \frac{5t}{16}}{1 - \frac{5t}{64}} \cdot \frac{1 + \frac{5t}{32}}{1 - \frac{5t}{4\cdot16}} \ldots
\]

We started with the step function on the values \( \frac{m}{1000} \) in the domain with \( m = 1 \) to \( m = 999 \). The number \( \frac{999}{1000} \) is the chosen maximal value \( \alpha \) and we chose to go all the way down to \( \frac{1}{1000} \) instead of \( \frac{1}{100} \). For every value \( p = \frac{m}{1000} \), for \( m = 1 \) to \( m = 999 \), a calculation was made for the expectation \( w_1(p) \). If \( \frac{m}{1000} \leq p < \frac{m+1}{1000} \), then \( w_0(p) \) was defined to be \( w_0(\frac{m}{1000}) \). For each case considered (as governed by the binomial expansion and the probability \( \hat{q} \)), if a \( q \) was used from some \( x^* \) coming toward some \( z_i \) (with \( x^* \) a distance of two from \( x \)), then we rounded down so that the value of \( w_0(q) \) corresponded to \( w_0(\frac{k}{1000}) \) where \( \frac{k}{1000} < q \leq \frac{k+1}{1000} \). We did not need to investigate any \( p \) with \( \frac{m}{1000} < p < \frac{m+1}{1000} \). The larger is \( p \) the higher goes \( w_1(p) \) and also the higher goes the ratio because \( w_0(p) \) is not increasing from \( \frac{m}{1000} \) to \( p \). So, to obtain a minimal growth ratio, it suffices to check only these 999 values for \( p \).

We followed the formula for calculating \( w_1 \) as presented in the previous section. Unfortunately, the function defined above through the generating function did not have a uniform positive growth rate, though for \( p \geq 0.055 \) the growth rate was positive and we discovered that 0.055 was the minimal value for \( p \) that was used to define \( w_1 \). We performed several iterations, moving from \( w_0 \) to \( w_1 \) to \( w_2 \), and so on to \( w_6 \), to get a step function that has a uniform growth rate for all choices of \( p = \frac{m}{1000} \). And then the resulting \( w_6 \) was redefined to be \( w_0 \).

Although the function \( w_0 \) is fictional, it is designed to be a plausible answer to the how the limit superior of the active pain behaves as a function of \( p \) relatively speaking, e.g. perhaps with a discount factor upwards or downwards, with the idea that in the limit the law of large numbers would stabilise almost all limit superiors.

No matter which \( p = \frac{m}{1000} \) was chosen, with \( 1 \leq m \leq 999 \), the ratio of \( w_1(p) \) to \( w_0(p) \) (defining \( w_0 \) by six iterations) is just slightly higher than 1.016. This means that the \( s \) from Proposition 2 is very close to 0.016.

We provide details of the lengthy calculations for those who are interested.

5 A Problem of \( \epsilon \)-Optimality and \( \epsilon \)-Stability

We can define the colouring rule in terms of a problem of local optimisation. At every point choices are made according to an objective function, which will be the sum total of three variables corresponding to the three types of choices that are made, the choice of five weights, the copying of those weights by adjacent points, and the choice of a passive pain level. We use the term solution for a function from \( X \) to \( C \) obeying the rule so as not to confuse it with "objective function".
The rule for the active colouring is already phrased in terms of an optimisation, the minimisation of active pain. As for the passive colourings, it is easy to make it the result of a minimisation. Let $c(y)$ be the sum total of weights directed at $y$. Choosing a level of $0 \leq b \leq 1$ at $y$ results in a cost of $b \cdot c(y) + (1 - b) \cdot (1 + r)$, with preference for $b = 0$ if $c(y) < 1 + r$, preference for $b = 1$ if $c(y) > 1 + r$, and any value for $b$ if $c(y) = 1 + r$. The copying of the weight of an adjacent point is done easily by taking the absolute value of the difference between the weight and the choice. Approximate copying will be done later in a linear way with finitely many options when we present the local optimisation again as a Bayesian game.

Of course for every positive $\epsilon$ there will be a measurable $\epsilon$-optimal solution where optimality is understood with respect to all the measurable options. On the other hand, given a measurable solution, we can integrate the objective function over the whole space and from the need for the weights inward to equal the weights outward it follows that expectation of the objective function will not go below $\frac{1}{9}(2^{-10} - \frac{2^{10}-1}{2^{10}}r)$ (just from the passive pain). This does not come close to the result almost everywhere when using some non-measurable solutions. But we are interested in another kind of $\epsilon$-optimality, which we call $\epsilon$-stability. For each $x \in X$ let $t(x)$ be the possible improvement in the objective function at $x$, keeping the solution for all other $y \neq x$ fixed. Let $\mu$ be a proper finitely additive extension. A $\mu$-measurable solution is $\epsilon$-stable if the $\mu$ expectation of $t(x)$ is no more than $\epsilon \geq 0$, meaning that there is no finite disjoint collection $A_1, \ldots, A_n$ of $\mu$ measurable sets such that the objection function can be improved by at least $t_i$ at all points in $A_i$ and $\sum_{i=1}^n \mu(A_i)t_i$ is greater than $\epsilon$. Another way of understanding $\epsilon$-stability is that $X$ is a uncountable space of human society or molecules, and the solution is $\epsilon$-stable if the gains from the individual deviations do not add up to an expectation of $\epsilon$. Our claim is that there is a positive $\epsilon$ such that no solution that is measurable with respect to any proper finitely additive extension is $\epsilon$-stable (and likewise for any $\epsilon^* < \epsilon$). This does not mean that if the deviations happened simultaneously there would be such an improvement for all concerned; indeed the result may be worse for all concerned.

There are two ways that a measurable solution must obey $\epsilon$-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. That can be formalised in the following way: if a solution is $\epsilon \cdot \delta$-stable, the subset where it diverges from optimality by more than $\delta$ cannot be of measure more than $\epsilon$.

In general, controlling for a potentially large deviation in a subset of size $\epsilon$ is the more difficult aspect to showing that there is no approximate stability. Nevertheless we start by considering the $\delta$ local optimality condition and build from it.

Up until now, we have not specified the positive value $r$, only that it was strictly between $0$ and $2^{-10}$. Now we fix $r$ to be exactly $2^{-11}$.

For any $\delta > 0$ the rule $Q_\delta$ is the same as $Q$ except that at all points $x$ obeying the rule $Q_\delta$ the colouring choices at $x$ are $\delta$-optimal and furthermore there is no terminating point $z$ of even level such that $x \to z$ and the weight given by $x$ to $z$ is more than $\frac{1}{12}2^{-12}$. This is a local rule, meaning that at each point there is no gain by more than $\delta$ from choosing a different colour, and there is no "large" contribution
of weight from a terminating point. By δ optimal we mean that an improvement by δ is allowed, but no more. In this way the rule becomes a closed relation and given a compact set of options where the δ optimisation can occur, the δ-optimal minimal solution is obtainable.

Again we introduce the concept of the w₀ process, except that the rule Q is replaced by the approximate rule Q^δ. As before, we know that after minimising in the w₀ process, the pain levels are less than what they were originally (the proof is identity). Using the w₀ process, we see again that the levels of pain implied by the Q^δ rule must be too high to allow for a substantial size of set where the passive pain is at least 1. There are two main differences however. First, we cannot make this claim for all positive passive pain levels, as we did for the Q rule. If that passive pain level is small compared to δ, one could slip away from the logic of the rule and allow for a substantially large set with that level of pain or less. Second, we have to re-introduce the influence of the terminating points, for the same reason, that extremely small pain levels could be involved. The w₀ process is defined on non-terminating points; the calculations of the w₀ process involve the probability of non-termination.

**Proposition 3.** Let w₀ and w₁ be functions defined on $\left[\frac{1}{100}, \alpha\right]$ satisfying the condition of Proposition 2 namely that there is some $s > 0$ such that for every $p$, $\log(w₁(p)) > \log(w₀(p)) + s$. For every $δ^* > 0$, define the function $w₁^δ(p)$ in the same way that $w₁$ is defined, however with the minimisation done with respect to the rule $Q^δ$ is obeyed instead of $Q$. We assume also that $w₀(\alpha) - w₀(\frac{1}{100}) > 1$. There is a δ small enough so that for every $δ^* < δ$, the function $w₁^δ(q)$ are defined entirely by the $w₀(q)$ with $q > \frac{1}{100}$ and, furthermore, $δ^* < δ$ implies that $w₁^δ(p) ≥ w₁(p) - δ^*\frac{1}{100}$ for all $p \in \left[\frac{1}{100}, \alpha\right]$.

**Proof.** First we consider deviation with respect to the passive pain level, and show that it is necessary that, with small enough $δ^*$, the summation of the weights directed toward the four $z_i$ with $x \rightarrow z_i$ must be at least 1. If the summation of the weights directed toward some $z_i$, not counting the terminating points, is less than $1 + r - 2^{-12} = 1 + 2^{-12}$, the choice of passive pain $b$ would be $b - \frac{1}{100} 2^{-12}$ away from the optimal choice of $b = 0$. It follows that $b$, the passive pain, cannot exceed $\frac{δ^*}{10 2^{-12}}$. As we assume that the active pain at $x$ by choosing $y$ is $p$, by the optimality assumption it cannot be that the active pain in the $z_i$ direction is less than $p - δ^*$. So we know that if $δ^*$ is less than $\frac{δ^*}{10 2^{-12}}$ then it must hold, after excluding the weights coming from terminating points, that the total weight directed toward $z_i$ is at least 1, a contradiction. (This is critical because the computer calculations are based on this sum being at least 1; if they were less than 1 the expected growth in values may not occur.)

Next, we show that in the process of minimising according to the rule $Q^δ$, no weight less than $\frac{1}{100}$ is used as long as $δ^* < 2^{-12}$. It means, from the previous paragraph, that the total weight coming from $x$ to $z_i$ is at least .91. As the weight from $x$ to $z_i$ is estimated at no less than $\frac{.91 - δ^*}{10}$, the passive pain at $z_i$ cannot exceed $\frac{1}{10}$ and the value of the product cannot exceed $\frac{1}{10}w₀(\frac{1}{100} + δ^*)$. On the other hand, the weight from $x$ to some $z_j$ must be at least 30, which will be multiplied by
at least \( w_0(0.09) \). By the monotonicity of \( w_0 \) we see that this is not a process of minimising according to the rule \( Q^{\delta^*} \).

Let \( \xi = w_0(\alpha) - w_0(\frac{1}{100}) \), and we assume that that \( \xi > 1 \) (as is true with the function used for the computer analysis). Now we show that for every \( p \in [\frac{1}{100}, \alpha] \) there is a uniform lower bound of \( \frac{1-p}{50p} \) to the minimal given by \( x \) to any \( z_i \) in the minimisation process using the rule \( Q^{\delta^*} \) with \( \delta^* < \frac{1-\alpha}{50\alpha} \) for some \( \delta > 0 \). As the only difference between \( w_1 \) and \( w_1^{\delta^*} \) pertains to the change in the passive pain levels at the \( z_i \) and that is determined by the reciprocals of the copying of the weights from \( x \) to the \( z_i \) (which cannot change by more than \( \delta^* \)), we get that the difference between the \( w_1 \) and \( w_1^{\delta^*} \) is bounded by \( \frac{50\delta^*}{1-\alpha} \). As \( \alpha \) is the largest probability for \( p \), the weight of at least \( 1-\alpha \) has to be distributed to the \( z_i \), \( i = 1, 2, 3, 4 \). If less than \( \frac{1-\alpha}{50\alpha} \) is given to some \( z_i \) then the maximal copying for that direction is \( \frac{3}{2} \frac{1-\alpha}{50\alpha} \) and the minimal passive pain level for \( z_i \) is \( \frac{2}{50\alpha} \frac{1-\alpha}{\alpha} - \delta^* \). With a minimal level of \( \frac{1}{100} \) for any weight sent to \( z_i \) other than that from \( x \) we get that the value of the product must be at least \( w_0(\frac{1}{100}) \left( \frac{3}{2} \frac{1-\alpha}{50\alpha} - \delta^* \right) \). On the other hand, there must be some \( z_j \) that receives at least \( \frac{1}{100} (1-\alpha) \) from \( x \), with a minimal copying for that direction of \( \frac{1-\alpha}{4} \). With the minimal passive pain level for \( z_i \) at \( \frac{4p}{1-\alpha} - \delta^* \) the value of the product is no more than \( w_0(\alpha) \left( \frac{4p}{1-\alpha} - \delta^* \right) \). It follows from \( \frac{w_0(\alpha)}{w_0(\frac{1}{100})} = \frac{\alpha}{\xi} \) that this cannot be minimisation according to the rule \( Q^{\delta^*} \).

In conclusion, \( \delta \) equal to \( \min(2^{-13}, \frac{1-\alpha}{50\alpha}) \) suffices. If the minimal is \( \frac{1-\alpha}{50\alpha} \) to the weight given to any \( z_i \), the change in passive pain level at any \( z_i \) between the \( Q \) rule evaluation and the \( Q^{\delta^*} \) evaluation when minimisation is according to the rule \( Q^{\delta^*} \) is no more than \( \delta^* \frac{1-\alpha}{50\alpha} \). The maximal difference in the product is \( \delta^* \frac{1-\alpha}{50\alpha} \). As minimisation according to the \( Q \) could occur with an even lower level elsewhere, the inequality concerning \( w_1^{\delta^*} \) and \( w_1 \) holds.

Now that we know that if \( \delta^* \) is small enough (and with local \( \delta^* \)-optimality a closed relation the minimalisation will again be unique, though that is not necessary), we can define a sub-martingale process as before, but with a growth rate slightly smaller than the \( s \). We can proceed to the main proof of this section, once we have proved two lemmas concerning terminating points.

**Lemma 5.** If \( y \) has passive pain of level \( v \) and \( x \) is a terminating point of odd level \( n \), and \( x \rightarrow y \) with \( x \) giving \( y \) weight of at least \( \frac{1}{10 \cdot 2^{\frac{n+1}{2}}} \) and each point between \( y \) and the terminating point of level 0 satisfies \( Q^{\delta} \) then \( \delta \geq v \left( \frac{2^{-12}}{10} \right)^{\frac{n+1}{2}} \).

**Proof.** Because \( x \) gives weight of at least \( \frac{1}{10 \cdot 2^{\frac{n+1}{2}}} \), its active pain is at least \( \frac{1}{10 \cdot 2^{\frac{n-1}{2}}} - \delta \) and therefore the terminating point of level \( n - 1 \) next to \( x \) has passive pain of at least \( \frac{1}{10 \cdot 2^{\frac{n+1}{2}}} (v \left( \frac{1}{10 \cdot 2^{\frac{n-1}{2}}} - \delta \right) - \delta \) (as the weight given to any other point cannot exceed \( 1 - \frac{1}{10 \cdot 2^{\frac{n+1}{2}}} \)). The result follows by induction.

**Lemma 6.** Let \( i \) be odd and let \( q_i \) be the probability of a chain \( x \rightarrow y \) having terminating level \( i \) (meaning that \( x \) is a terminating point of level \( i \)). Then the probability \( q_i \) is less than \( \frac{1}{128} \) and the probability of \( q_{i-1} \) is less than \( \frac{1}{128} \left( \frac{1}{6} \right)^{\frac{i+1}{2}} \).

**Proof.** The probability that \( x \rightarrow z \) and \( z \neq y \) is terminating of level 0 is exactly \( 2^{-9} \). Since there are four such \( z \), \( q_1 \) is no more than \( 4 \times 2^{-9} \). Now assume that \( x \rightarrow z \)
and $z \neq y$ is a terminating point of level $i - 1$. There is at least one $x^*$ that is a terminating point of level $i - 2$ with $x^* \to z$ and no other $\hat{x} \to z$ that is non terminating. The probability is no more than $9 \cdot (1 - \frac{1}{2})^8 q_{i-2}$, where $\hat{q}$ is approximately .991. Since this could happen in any one of four places, $q_i$ is no more than $4 \cdot 9 \cdot (1 - \frac{1}{2})^8 < \frac{1}{6} \blacksquare$

**Theorem 2.** There is a positive $\epsilon$ small enough so that there is no $\epsilon$-stable solution to the rule $Q$ that is measurable with respect to any proper finitely additive extension.

**Proof.** We will prove that, with sufficiently small positive $\epsilon$ and $\delta^*$, the subset where $Q^{\delta^*}$ does not hold must exceed $\epsilon$, given that the solution is measurable.

From Proposition 3, there is some $B > 0$ such that if $\delta$ is small enough the minimisation process according to the rule $Q^\delta$ on the $w_0$ process and subtraction at each level by $\frac{1}{2}$ defines a logarithmic sub-martingale process such that the variance at each stage does not exceed $B > 0$. Small enough means that $\delta$ is smaller than the passive pain level at every stage times the $\delta$ given at the conclusion of Proposition 3. Also from Proposition 3 we have a bound $M > 1$ on the proportion that the passive pain level can drop at each stage (also determined by the minimal weight condition from the proof of Proposition 3). By the Chebyshev martingale inequality (using the same arguments as before) there is a positive integer $N$ such that after a distance of $N$ from any starting vertex $y$ on the graph of $X$ where the passive pain level is at least $\frac{1}{2}$, the probability is at least $1 - \frac{1}{10}2^{-12}$ that some point in the resulting chain up to the stage of $N$ (following the rule $Q^\delta$) must be given a passive pain level greater than 2, (which is not possible). We observe that the same arguments behind Lemma 3 and Lemma 4 apply to the rule $Q^\delta$. The number of vertices of distance $N$ away from a point in a chain of length $N$ does not exceed $50^N$. So, by making $2\epsilon$ smaller than $\frac{50^{-12}}{50^N}$ and $\delta$ smaller than $\frac{1}{2}e^{-NM}\delta$ (where $\delta$ is the quantity determined in Proposition 3), we guarantee that if a set of size no more than $2\epsilon$ is excluded from the rule $Q^\delta$ then $\epsilon\delta$-stability is a contradiction to proper finitely additive measurability.

Starting from any $x \to y$ where the passive pain at $y$ is at least 1, we have the sufficient rarity of such points by the pain levels in the original process exceeding those of the $w_0$ process.

But we still have to remove all terminating points of odd level putting in a weight of at least $\frac{1}{10}2^{-12}$ into the non-terminating chain generated by $x \to y$. By Lemma 6 there is some odd $i$ such that the chances of some terminating point of level $i$ or more sending more than $\frac{1}{10}2^{-12}$ to any of these vertices, at most $50^N$ of them, is less than $\epsilon$. So, we decrease our $\delta^*$ to $(\frac{2^{-12}}{10})^{\frac{1}{2}}\delta$ to have our pair $\delta^*$ and $\epsilon$.

$\blacksquare$

**A Bayesian Game**

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. R. Simon [S] 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. R. Simon and G. Tomkowicz [4] showed that there is a Bayesian game with non-measurable equilibria but no Borel measurable \( \epsilon \)-equilibrium for small enough positive \( \epsilon \). That construction involved the action of a non-amenable semi-group.

Bayesian games present a novel aspect to optimisation and stability, because the knowledge of a player is defined by a partition of the space and with finitely many players there are finitely different partitions. We could perceive optimisation as the optimisation of one player, but that perspective we showed already to be not very interesting, as that is nothing more than noticing that one can do much better with non-measurable solutions than with measurable ones. With our previous concept of \( \epsilon \)-stability, we could perceive a game with uncountably many players, as described above. But with Bayesian games with finitely many players, because knowledge is defined by finitely many partitions, local deviation of a player does not get in the way of global optimisation for that player, given of course that the strategies of the other players are fixed. With Bayesian games we can consider deviation equivalently in global and local terms, and the information of the players, defined by their partitions, gives added depth to the structure. In [S] and [4] we defined a Harsanyi \( \epsilon \)-equilibrium for a positive \( \epsilon \): all players in the \( \epsilon \)-equilibrium choose measurable strategies and there is no Borel measurable deviation by some player to another measurable strategy resulting in an expected gain of more than \( \epsilon \).

This definition of Harsanyi equilibria can be extended to proper finitely additive extensions. If the local beliefs of the players are defined by certain ergodic operators, then those ergodic operators must remain measure invariant with respect to the finitely additive extension. For example, if at a point \( x \) the player believes that \( x \) and \( T(x) \) are equally likely for some ergodic operator \( T \) (and the same is believed at \( T(x) \)), then they are given equal probability by this player. It is valid to say that this belief could not be violated in any finitely additive extension. This is especially true of the following re-working of the \( Q \) rule in terms of a Bayesian game.

A Bayesian \( \epsilon \)-equilibrium is a way for each player to choose local strategies, not necessarily in a measurable way, so that evaluated at each location the player does not gain more than \( \epsilon \) by choosing another strategy. Hellman showed that there was a two person Bayesian game without Bayesian \( \epsilon \)-equilibria for sufficient small positive \( \epsilon \). For a positive \( \epsilon \), it is more difficult to find a game that does not have a Harsanyi \( \epsilon \)-equilibrium. This is because a Harsanyi \( \epsilon \)-equilibrium could employ a very small set (for example of measure less than \( \epsilon \)) where the deviation from local \( \epsilon \)-equilibrium is significant.

Not only does the following Bayesian game fail to have \( \epsilon \)-Harsanyi equilibria for sufficiently small \( \epsilon \) with respect to all proper finitely additive extensions, this game can be constructed using only two players. Indeed this example answers all the questions posed in the conclusion of [4]. It is also an ergodic game [S].

Let \( G \) be the group generated freely by five generators, \( T_1, T_2, T_3, T_4, T_5 \) and let \( X \) be the Cantor set \( \{-1, 1\}^G \). Let \( A \) be the set \( \{a_i^+, a_i^- \mid i = 1, 2, 3, 4, 5\} \) of cardinality 10 and let \( B \) be the set \( \{b_i^+, b_i^- \mid i = 1, 2, 3, 4, 5\} \) of cardinality 10. We assume that \( A \) and \( B \) are disjoint. Let \( C \) be the set \( A \cup B \) of cardinality 20 and let \( \Omega \) be the space \( X \times C \). Give the set \( X \) the canonical probability distribution \( m \).
such that the measure of a cylinder set defined by
\[ \{x \mid x^g_1 = f_1, \ldots, x^g_i = f_i \} \]
is equal to \(2^{-l}\) for every sequence \(f_1, \ldots, f_i\) of choices in \([-1, 1]\) and \(g_1, \ldots, g_i\) are mutually distinct. Define the canonical Borel measure \(\mu\) on \(\Omega\) by
\[ \mu(A \times \{c\}) = \frac{m(A)}{2^0}, \]
for every Borel measurable set \(A\) in \(X\) and any choice of \(c\) in \(C\).

There are two players, the active player, called the green player, and the passive player, called the red player. For every \(x \in X\), the green player has the information set
\[ \{(x \times A) \cup \bigcup_{i=1,2,3,4,5} \{(T_i(x), b_i^-), (T_i^{-1}(x), b_i^+)\}\}. \]
For every \(y \in X\), the red player has the information set
\[ \{(y \times B) \cup \bigcup_{i=1,2,3,4,5} \{(T_i(y), a_i^-), T_i^{-1}(y), a_i^+)\}\}. \]
Notice that each information set is of cardinality 20 and for both players these sets partition the space. To identify the information set of the player, the green player is at \(x\) if \(\{x\} \times A\) is half of its information set and the red player is at \(y\) if \(\{y\} \times B\) is half of its information set. We will also refer to \((x, a_i^+)\) as \((x, a_y)\) where \(y = T_i(x)\), \((x, a_i^-)\) as \((x, a_y)\) where \(y = T_i^{-1}(x)\), \((y, b_i^+)\) as \((y, b_x)\) where \(x = T_i^{-1}(y)\), and \((y, b_i^-)\) as \((y, b_x)\) where \(x = T_i(y)\).

The green player has the choice of 5 actions, \(t_1, t_2, t_3, t_4, t_5\). A strategy for the green player at any \(x\) is a point in the four dimensional simplex \(\Delta(t_1, t_2, t_3, t_4, t_5)\).

The red player at \(y\) has the choice of \(2 \cdot M^{d(y)}\) actions, where \(M\) is a very large positive integer, size to be determined later. The set of actions is
\[ \{c, u\} \times \prod_{x \in S(y)} \{m_x \mid 0 \leq m_x \leq M - 1\}. \]
The symbol \(c\) stands for “crowded” and \(u\) for “uncrowded”. The choice of a mixed strategy for the red player is for some point in the \(2 \cdot M^{d(y)} - 1\) dimensional simplex.

The payoffs for the green player at \(x\) take place only in \(\{x\} \times A\), meaning that in the other ten locations the payoff is uniformly zero. The payoffs for the red player at \(y\) take place only in \(\{y\} \times B\). It is more restrictive than this. The payoffs for the green player at \(x\) take place only in the five locations \(\{(x, a_i^+) \mid i = 1, 2, 4, 5\}\) if \(x^e = +1\) or only in the five locations \(\{(x, a_i^-) \mid i = 1, 2, 4, 5\}\) if \(x^e = -1\). The payoffs for the red player at \(y\) take place only in that subset of \(B\) corresponding to the subset \(S(y)\) (meaning only at the \(b_x\) with \(x \in S(y)\)). With both players, as each gives the probability \(\frac{1}{20}\) to each point in its information set, the payoff is determined by summing over all the points giving equal weight to each. The key to understanding is that whatever is played by the green player \(x\) is done uniformly throughout its information set \(\{x\} \times A \cup \bigcup_{i=1,2,3,4,5} \{(T_i(x), b_i^-), (T_i^{-1}(x), b_i^+)\}\), and the same is true for the red player at \(y\) and its information set.
First we define the payoffs for the green player. We consider what happens to the green player at \( x \) when choosing the action \( t_i \). The action \( t_i \) has a payoff consequence only at the point \((x, a_y)\) where \( y = T_i^{x^*}\). Given that the red player at \( y \) chooses \((c, m_{x^*}, \ast)\), where \( \ast \) stands for any choices of \( m_{x^*} \) for other \( x' \in S(y) \), the payoff to the green player at \((x, a_y)\) is \(-\frac{m_y}{M}\). Otherwise for all combination with \( u \) instead of \( c \) the the payoff is 0.

Now we define the payoffs for the red player. For any \( x \in S(y) \), let \( y = T_i^{x^*} \) and let \( t_y \) be the action \( t_i \). For any \( z = T_i^{x^*} \), with \( j \neq i \), let \( t_z \) be that action \( t_j \). First consider a piece-wise linear convex function \( f : [0, 1] \rightarrow \mathbb{R} \), where \( f = \max_k f_k \) for some affine functions \( f_0, \ldots, f_M-1 \) where \( f \) is equal to \( f_k \) on \([\frac{k}{M}, \frac{k+1}{M}]\). Let \( s_j^+ \) and \( s_j^- \) be defined by \( f_k(0) = s_k^- \) and \( f_k(1) = s_k^+ \), and the difference in slopes between consecutive \( f_i \) and \( f_{i+1} \) is always at least 1. Define the value of the actions \((c, m_x, \ast)\) played against \( t_y \) at \((y, b_x)\) to be \( s_{m_x}^+ + 1 \), the value of the actions \((u, m_x, \ast)\) played against \( t_y \) at \((y, b_x)\) to be \( s_{m_x}^- + 1 + r \), the value of the actions \((c, m_x, \ast)\) played against \( t_z \) at \((y, b_x)\) to be \( s_{m_x}^- \), and the value of the actions \((u, m_x, \ast)\) played against \( t_z \) at \((y, b_x)\) to be \( s_{m_x}^- + 1 + r \) (for \( z \neq x \)).

Because the consequence for the red player at \( y \) by choosing some \( m_x \) for \( x \in S(y) \) lies entirely at the point \((y, b_x)\) and is also independent of the choice for \( c \) or \( u \), the red player will chose the marginal probabilities for \( m_x \) according to \( s_{m_x}^+ \) and \( s_{m_x}^- \). By the structure of those values, no more than two \( m_x \) will be chosen in equilibrium, and only two adjacent \( m_x - 1 \) and \( m_x \) if the probability for \( t_y \) is exactly \( \frac{m_x}{M} \). When the probability for \( t_y \) lies strictly between \( \frac{m_x}{M} \) and \( \frac{m_x + 1}{M} \) then only \( m_x \) will be chosen in equilibrium.

Notice that in equilibrium this game not only approximates the colouring rule of the previous sections, and it can be done so in a way for which the computer calculations also apply. If the green player at \( x \) chooses the action \( t_y \) with probability \( q \), the red player at \( y \) will respond only with various combinations of \((c, m_{[qM]}, \ast), (u, m_{[qM]}, \ast)\) and possibly with some \((c, m_{[qM-1]}, \ast), (u, m_{[qM-1]}, \ast)\) if \( qM \) is an integer. The payoff for the green player at \( x \) will be the total probability at \( y \) for playing \( c \) times some quantity that is between \([q_1M - 1] \) and \([q_1M]\). As the payoff for the green player at \( y \) is the \( \frac{1}{2} \) of the sum of the payoffs at the \((y, S(y))\), there will be only a choice for \( c \) if the sum of the probabilities for the \((t_x \mid x \in S(y))\) is greater than \( 1 + r \), only a choice for \( u \) if that sum is less than \( 1 + r \), and any mixed use of \( c \) and \( u \) if that sum is exactly \( 1 + r \).

To show a lack of a Harsanyi-\( \epsilon \)-equilibrium for any proper finitely additive extension, using our previous argument for the lack of an \( \epsilon \)-stable solution, we require that the process of copying weights is done with proper precision. Whatever \( \delta \) worked for \( \epsilon \delta \)-stability argument (using the Chebyshev martingale inequality), we divide by 3 and declare this to be the quantity needed for the lack of finitely additive measurable \( \epsilon \delta /3 \)-equilibria for this Bayesian game. We make \( M \) be larger than \( \frac{2}{\delta} \) to insure that there is no inaccuracy up to \( \frac{\delta}{3} \) resulting from the intervals used. But lastly, we need to know that there is no relevant distortion from the mixture of the \( c \) and the \( u \) coordinates with the occasional choice of a level \( m_j \) that is not a good copy of the actual weight sent from the relevant point. We need to know that the summation of the probabilities given to the actions \((c, m_j, \ast)\) is sufficiently close to the average value for \( m_j \) times the average proportion for \( c \) (the product
of expectations from the marginals). This is because we need to use the average proportion for $c$ to be the basis for the passive pain defined independently of the various neighbouring vertices sending in weights. Let’s suppose that the level $m_j$ is incorrect when $m_i$ is the choice closest to the correct choice on the same side as $m_j$. Due to the slopes of the lines defining the payoffs, we know that the cost of this mistake is at least $|j-i|(|j-i|-1)/2 q_j$ where $q_j$ is the probability of using $m_j$. We have that the summation over $j$ of the $|j-i|(|j-i|-1)/2 q_j$ cannot exceed $\delta/3$. It follows that $\frac{1}{M} \sum_j q_j |i-j|$ cannot exceed $2\delta/M$. By choosing $M$ greater than $\frac{4}{\delta}$, we have the needed accuracy.

6 Conclusion

We demonstrated a way to satisfy the colouring rule that induces no pain at all. As it involves giving full weight in some direction from every point, and no point receives weight more than 1, by a theorem of [5], a $G$ paradoxical decomposition in the space exists in the sigma algebra generated by the inverse images of the five colours, the Borel sets and the operations in $G$. A natural question is the following:

**Question 1.** For all colourings satisfying the rule $Q$ is there a finite partition of the colour space into Borel subsets so that a $G$ paradoxical decomposition of the space exists in the sigma algebra generated by the inverse images of the finite partition, the Borel sets and the actions in $G$? Is this true in general for all paradoxical colouring rules when there is a compact and convex set of colours?

A further issue is raised by the expected value of the non-measurable solutions. We know about the rule $Q$ that there exists non-measurable solutions where optimality is perfect, meaning the universal cost or pain of 0. And we know that with measurable solutions there is an average passive pain level above $\frac{2-11}{9}$. By slight manipulation of the cost function, we could create a new kind of paradox:

**Question 2.** Does there exist a problem of local optimisation or a Bayesian game such that the optimisation can be accomplished locally or the values can be measured globally, but not both simultaneously?

Structuring a Bayesian game with non-invertible ergodic operators makes it easy to demonstrate that there is a Bayesian game with no Harsanyi measurable $\epsilon$-equilibrium [4], but we were not able to create a probabilistic paradoxical rule with non-invertible ergodic operators used to define the colouring rule. Our rule $Q$ uses a lot of back and forth causation, which is problematic with non-invertible operators. In [5], the measure preserving transformations used to define the colouring rule were called descendents.

**Question 3.** Does there exist a probabilistic colouring rule with some descendents defined by non-invertible measure preserving transformations?

Theorem [1] uses a free non-abelian group of rank 5. Given the existence of non-amenable groups without free non-abelian subgroups, demonstrated by Olshanskii and Grigorchuk, (see [TW], Chapter 12 for the details) it is natural to ask the following:
Question 4. *Does there exist a probabilistic paradoxical colouring rule that uses a non-amenable group without free non-abelian subgroups?*

The idea behind Question 4 is related to the complexity behind the proof of Theorem 1. Recall that two or three free choices were not enough to obtain a paradoxical rule. So it is natural to investigate and describe if the required complexity can be forced by generators that are not independent.

The proof of Theorem 2 seems convoluted. Terminating points and non-terminating points are treated separately, and it would be nice to have a unified approach. The problem is that in the calculations behind Theorem 1 integrating the effect of terminating points into the argument would involve a division by 0 (as we divide by one less than the degree of the vertex). Indeed terminating points are such that they need infinite levels of pain in order to avoid sending all weight toward them. The present approach is not efficient for establishing a good upper bound for the $\epsilon$ for which there is no measurable $\epsilon$-optimal or $\epsilon$-stable solution.

Question 5. *What is the largest positive $\epsilon$ such that there is a probabilistic paradoxical colouring rule defined by a local optimisation where the objective function is between 0 and 1 and there is no $\epsilon$-stable solution that is measurable with respect to any proper finitely additive extension?*

References

[1] Z. Hellman, *A Game with no Bayesian Approximate Equilibria*, Journal of Economic Theory, 153(C) (2014), pp. 138-151.

[2] J. Mycielski and G. Tomkowicz, *Shadows of the Axiom of Choice in the universe $L(R)$*, Arch. Math. Logic, 57 (2018), pp. 607-616.

[3] R. S. Simon, *Games of Incomplete Information, Ergodic Theory, and the Measurability of Equilibria*, Israel J. Math., 138, 1, (2003) pp. 73-92.

[4] R. S. Simon and G. Tomkowicz, *A Bayesian Game without $\epsilon$-equilibria*, Israel J. of Math. 227 (2018), 215-231.

[5] R. S. Simon and G. Tomkowicz, *Paradoxical decompositions and finitary colouring rules*, Preprint.

[6] G. Tomkowicz and S. Wagon, *The Banach-Tarski Paradox, Second Edition*, Cambridge University Press, 2016.