Electronic Journal of Probability

Electron. J. Probab. **28** (2023), article no. 143, 1–60.
ISSN: 1083-6489 https://doi.org/10.1214/23-EJP1046

On a class of PCA with size-3 neighborhood and their applications in percolation games*

Dhruv Bhasin† Sayar Karmakar‡ Moumanti Podder§
Souvik Roy¶

Abstract

Different versions of percolation games on $\mathbb{Z}^2$, with parameters p and q that indicate, respectively, the probability with which a site in $\mathbb{Z}^2$ is labeled a trap and the probability with which it is labeled a target, are shown to have probability 0 of culminating in draws when $p + q > 0$. We show that, for fixed p and q, the probability of draw in each of these games is 0 if and only if a certain 1-dimensional probabilistic cellular automaton (PCA) $F_{p,q}$ with a size-3 neighborhood is ergodic. This allows us to conclude that $F_{p,q}$ is ergodic whenever $p + q > 0$, thereby rigorously establishing ergodicity for a considerable class of PCAs that tie in closely with important topics such as the enumeration of directed animals, broadcasting of information on directed infinite lattices, examining reliability of computations against the presence of noise etc. The key to our proof is the technique of *weight functions*. We include extensive discussions on *game theoretic* PCAs to which this technique may be applicable to establish ergodicity, and on percolation games to which this technique may be applicable to explore the 'regimes' (depending on the underlying parameter(s), such as (p, q) in our case) in which the probabilities of draw are 0.

Keywords: percolation games on lattices; two-player combinatorial games; probabilistic cellular automata; ergodicity; probability of draw; weight function; potential function.
MSC2020 subject classifications: 05C57; 37B15; 37A25; 68Q80.
Submitted to EJP on November 9, 2022, final version accepted on October 21, 2023.

---

*For the duration of this project, Dhruv Bhasin has been supported by Grant No. 0203/2/2021/RD-II/3033 awarded by the National Board for Higher Mathematics (NBHM), Sayar Karmakar has been supported by Grant No. NSF DMS 2124222 awarded by the National Science Foundation (NSF), and Moumanti Podder has been supported by Grant No. SERB CRG/2021/006785 awarded by the Science and Engineering Research Board (SERB). In addition to these funders, the authors express their gratitude towards their respective institutes / university for supporting them in myriad ways and for fostering wonderful environments conducive to collaborations in research.

†IISER Pune. E-mail: bhasin.dhruv@students.iiserpune.ac.in
‡University of Florida. E-mail: sayarkarmakar@ufl.edu
§IISER Pune. E-mail: moumanti@iiserpune.ac.in
¶ISI Kolkata. E-mail: souvik.2004@gmail.com

1 Introduction

1.1 Overview of the paper

There are two aspects to this paper that share a deep connection with each other: a class of 1-dimensional *probabilistic cellular automata* (henceforth abbreviated as PCA), and a class of *percolation games* that is studied on the 2-dimensional square lattice $\mathbb{Z}^2$. At the very outset, we give a brief sketch of the percolation games and the specific PCAs that we focus on, while their detailed descriptions are provided in §2.1 and §2.2 respectively.

The key components constituting our PCA (that we henceforth refer to as $F_{p,q}$) are the *alphabet* $\{0,1\}$, the *universe* $\mathbb{Z}$, the *neighborhood* $\mathcal{N} = \{0,1,2\}$, and the parameters p and q with $(p,q) \in \mathcal{S}$, where

$$\mathcal{S} = \{(p',q') \in [0,1]^2 : 0 < p' + q' \leqslant 1\}. \tag{1.1}$$

A *configuration* for this PCA at a given time t is an element η_t of $\{0,1\}^{\mathbb{Z}}$, and its evolution can be described as follows: the state $\eta_{t+1}(n)$ of a site $n \in \mathbb{Z}$ at time $t+1$ is updated, independent of the updates happening at all other sites, according to a probability distribution that is supported on $\{0,1\}$ and that depends on $(\eta_t(n+i) : i \in \mathcal{N})$, i.e. on the states $\eta_t(n)$, $\eta_t(n+1)$ and $\eta_t(n+2)$ of the sites n, $n+1$ and $n+2$, respectively, at time t. The precise description of these probability distributions can be found in §2.2. We remark here that in the sequel, much of our mathematical analysis is carried out on a different but related PCA, $\widehat{\mathrm{F}}_{p,q}$, with alphabet $\{0,?,1\}$, that is referred to as the *envelope* to $F_{p,q}$, since it serves to provide a more direct connection with the percolation game we study in this paper (see §2.3).

We now come to an overview of the percolation game we are concerned with, in which the parameters p and q, with $(p,q) \in \mathcal{S}$, are fixed *a priori*. We begin by assigning, independently to every vertex of $\mathbb{Z}^2$, a (random) label that reads *trap* with probability p, *target* with probability q, and *open* with the remaining probability $r = 1-p-q$. A token is placed at a vertex of $\mathbb{Z}^2$, termed the *initial vertex*, at the start of the game, and two players take turns to make moves. A *move* involves relocating the token from its current position (x,y) (say) to any vertex in the set $\mathrm{Out}(x,y)$, where

$$\mathrm{Out}(x,y) = \{(x,y+1),(x+1,y+1),(x+2,y+1)\}. \tag{1.2}$$

A player loses if she is forced to relocate the token to a vertex labeled a trap, and wins if she is able to relocate the token to a vertex labeled a target. The game may also continue forever, with neither player being able to reach a target nor being able to force her opponent to fall into a trap, and in this case, we say that the game results in a draw.

The primary goal of this paper is to prove Theorem 1.1, stated below, the first assertion of which establishes ergodicity (see Definition 2.1) for the PCA $F_{p,q}$, while the second concerns itself with the outcome of draw in the percolation game described above.

Theorem 1.1. *For all $(p,q) \in \mathcal{S}$, the PCA $F_{p,q}$ is ergodic, and the probability of draw in the percolation game described above is 0.*

We give a brief outline of the essential constituents of the proof of this theorem. Recall the envelope PCA $\widehat{\mathrm{F}}_{p,q}$ corresponding to $F_{p,q}$, with alphabet $\{0,?,1\}$, alluded to above. For each $(p,q) \in \mathcal{S}$, it is shown, via connections explained in §2.3, that the probability of the event that the percolation game starting from the origin in $\mathbb{Z}^2$ culminates in a draw is equal to the probability that the symbol $?$ occupies the origin in $\mathbb{Z}$ under a certain stationary distribution (see Definition 2.1) for $\widehat{\mathrm{F}}_{p,q}$. This stationary distribution for $\widehat{\mathrm{F}}_{p,q}$ is the same as the probability distribution of winning (for the first player), losing (for the first player) and draw positions (indicated, respectively, by the states W, L and D that are introduced at the beginning of §2.3) on *any* horizontal line $H_k = \{(x, k) : x \in \mathbb{Z}\}$ of $\mathbb{Z}^2$, for our percolation game. Next, Theorem 2.4 is used to establish that the probability of appearance of $?$ at any given site in $\mathbb{Z}$ is 0 under every stationary distribution for $\widehat{\mathrm{F}}_{p,q}$, for each $(p, q) \in \mathcal{S}$, leading to the conclusion that the probability of draw in our percolation game is 0 for each $(p, q) \in \mathcal{S}$. Finally, Proposition 2.2 guarantees that the probability of draw in our percolation game is 0 if and only if the PCA $F_{p,q}$ is ergodic, yielding the conclusion that $F_{p,q}$ is ergodic for each $(p, q) \in \mathcal{S}$.

The key tool used in the proof of Theorem 2.4 is the technique of *weight functions*. In addition to proving the main result, Theorem 1.1, of this paper, we include an extensive discussion on the class of percolation games to which this technique may be applicable, with possibly suitable modifications, to explore the 'regimes' (depending on the underlying parameters, such as (p, q) in our case), referred to as *critical regions*, in which such games have probability 0 of culminating in draws. We also provide insight into the class of *game theoretic* PCAs to which the technique of weight functions may be applicable in order to establish ergodicity properties, or to gain an understanding of the condition(s) to be imposed on the underlying parameter(s) of these PCAs that guarantee ergodicity. We also provide a formal game theoretic formulation of our problem, following which we establish connections between our work and the existing literature, speculate on the various directions in which future research can be carried out (for instance, by generalizing the set of actions, considering arbitrary mover-sequences, generalizing the event of draw itself, exploring possible monotonicity properties of the probabilities of draw, studying percolation games on higher dimensions, studying the values of such games etc.).

1.2 Brief discussion of the literature

A *cellular automaton*, also referred to as a *deterministic cellular automaton* (and henceforth abbreviated as CA) is a discrete dynamical system that consists of

1. a regular network of automata (or *cells*) indexed by $\mathbb{Z}^d$ for some $d \in \mathbb{N}$,
2. a finite set of states $\mathcal{A}$ termed the *alphabet*,
3. a finite set of indices $\mathcal{N} \subset \mathbb{Z}^d$ termed the *neighborhood*,
4. and a *local update rule* $f : \mathcal{A}^{\mathcal{N}} \to \mathcal{A}$.

The state (in $\mathcal{A}$) of each cell $\mathbf{x} \in \mathbb{Z}^d$ is updated by applying f to the current states of all the cells $\mathbf{x} + \mathbf{y}$ where $\mathbf{y} \in \mathcal{N}$, and this process is repeated at discrete time steps (see [26] for a detailed survey on CAs). It is not surprising, therefore, that many naturally occurring processes (such as collisions between moving point particles in a regular lattice that serves as a model for fluid dynamics) that are governed by local and homogeneous underlying rules can be efficiently modeled and simulated using CAs. CAs correspond to all those functions on $\mathcal{A}^{\mathbb{Z}^d}$ that are continuous with respect to the product topology and that commute with translations (see Corollary 76, Proposition 77 and other relevant discussions in Section 5 of [27], and [24]). Moreover, CAs serve as mathematical models for massive parallel computations, their relatively simple update rules allow them to be computationally universal, and they are capable of simulating any Turing machine ([20, 42, 14, 43, 37]). It is no wonder, consequently, that CAs have emerged as a useful tool for studying computations in natural processes as well as physical aspects and physical limits of computation.

A *probabilistic cellular automaton* (PCA) is obtained when the corresponding update rule is random. PCAs may be interpreted as discrete-time Markov chains on the state space $\mathcal{A}^{\mathbb{Z}^d}$ as well as generalizations of CAs. In a d-dimensional PCA, the state $\eta_{t+1}(\mathbf{x})$, at time $t+1$, of a cell $\mathbf{x} \in \mathbb{Z}^d$ is a random variable whose probability distribution (supported on $\mathcal{A}$) is governed by the states $\eta_t(\mathbf{x}+\mathbf{y})$, at time t, of the cells $\mathbf{x}+\mathbf{y}$, for all $\mathbf{y} \in \mathcal{N}$. PCAs are obtained in computer science via *random perturbations* of CAs, and the motivation behind studying them includes

1. investigating the fault-tolerant computational capability of CAs ([35, 36]),
2. classifying elementary CAs by using their robustness to errors and perturbations as a discriminating criterion ([31, 7]),
3. the intimate connection PCAs have with Gibbs potentials and Gibbs measures in statistical mechanics ([40, 30, 21]) as well as with combinatorial models such as directed animals ([33, 6, 8, 1, 12]), queues ([33], §4.3) and percolation ([23, 22, 29, 39, 41]),
4. and the crucial role PCAs play in modeling several complex systems appearing in physics, chemistry and biology.

We refer the reader to [13], and the more recent [33], for a detailed survey on how the theory of PCAs has developed over the years.

In addition, to give the reader a glimpse into the vast and variegated applications of PCAs in various disciplines, we refer to [32] that offers an insight into how PCAs are implemented in probability, statistical mechanics, computer sciences, natural sciences, dynamical systems and computational cell biology, to [30], [19], [44], [21] and [3] that discuss the implementations of PCAs in statistical mechanics and dynamical systems, to [45] that applies PCAs as a quantitative framework in astrobiology, to [15] that studies DNA sequence evolutions and cellular mutations in computational cellular biology using PCAs, and to [38] as an instance of how PCAs find applications in chemistry. We reiterate to the reader that the references mentioned above form just a minuscule fraction of the diverse ways PCAs have found applications in several branches of mathematics, physics, chemistry and biological sciences.

We now come to a more nuanced discussion, focusing on a few of those articles (keeping in mind that this list too is, by no means, exhaustive) that study PCAs from perspectives closely related to how we approach $F_{p,q}$ and $\widehat{\mathrm{F}}_{p,q}$. We begin our discussion with [25], not only because this serves as the primary inspiration for our work in this paper, but also because percolation games were introduced and studied for the first time in [25]. The percolation game considered in [25] admits $\mathrm{Out}(x,y) = \{(x+1,y),(x,y+1)\}$ for every $(x,y) \in \mathbb{Z}^2$, and whether the probability of a draw is strictly positive or not is shown to be intimately connected with the ergodicity of a family of elementary 1-dimensional PCAs (see §2.2 for details). On the other hand, certain analogous games played on various directed graphs in higher dimensions (such as on an oriented version of the even sub-lattice of $\mathbb{Z}^d$ for all $d \geqslant 3$) are shown to exhibit positive probabilities of draw for suitable values of the parameters involved – a fact that is established via dimension reduction to a hard-core lattice gas model in dimension $d-1$, and by showing that draws occur whenever the corresponding hard-core model fails to have a unique Gibbs measure.

In [34], it is shown that various families of CAs, such as nilpotent CAs, permutive CAs, gliders, CAs with spreading symbols, surjective CAs and algebraic CAs, are highly unstable against various types of noises, such as zero-range noise, memoryless noise, additive noise, permutive noise and birth-death noise, in the sense that they forget their initial conditions under the slightest positive noise – a fact that is reflected in the ergodicity of the PCAs that result from these CAs under the application of said noise. [34] also discusses results on the stronger notion of uniform ergodicity of PCAs as well as spatial mixing, computability and admittance of perfect sampling algorithms for their unique stationary or invariant measures, and the techniques used include couplings, entropy and Fourier analysis. In [11], a family of 1-dimensional PCAs with memory two is studied as these arise naturally from the 8-vertex model, directed animals, gaz models, TASEP and various other models of statistical physics – in such a PCA, the state $\eta_{t+1}(n)$ of the site $n \in \mathbb{Z}$ at time $t+1$ is a random variable whose probability distribution is a function of the states $\eta_t(n-1)$ and $\eta_t(n+1)$ of its two nearest neighbors $n-1$ and $n+1$ at time t, and its own state $\eta_{t-1}(n)$ at time $t-1$. In [11], conditions are proposed under which the invariant measures for these PCAs can be expressed either in a product form or in a Markovian form, ergodicity results that hold in this context are proved, and the phenomenon of reversibility of the stationary space-time diagrams of these PCAs is investigated, leading to the discovery of families of Gibbs random fields on the square lattice that have fascinating geometric and combinatorial properties. In [9], a computer-assisted proof of ergodicity is provided for two PCAs whose update rules can be respectively expressed as $\eta_{t+1}(n) = \mathrm{BSC}_p(\mathrm{NAND}(\eta_t(n-1), \eta_t(n)))$ (referred to as the vertex noise scenario) and $\eta_{t+1}(n) = \mathrm{NAND}(\mathrm{BSC}_p(\eta_t(n-1), \eta_t(n)))$ (referred to as the edge noise scenario), with $p \in (0, \epsilon)$ for some suitable $\epsilon > 0$. Here, BSC_p refers to a binary symmetric channel that takes a bit as input and flips it with probability p, leaving it unchanged with probability $1-p$. Similar to the approach adopted in this paper, [9] utilizes the notion of *weight function* or *potential function* (see §4 for details) that was introduced in [25], but instead of any explicit manual computations, they implement local feasibility of a suitable polynomial linear programming (PLP) to guarantee the existence of a desired potential function that then helps establish the above-mentioned ergodicity results.

1.3 Organization of the paper

The percolation game we investigate in this paper is discussed and illustrated in §2.1, the PCA $F_{p,q}$ is formally introduced in §2.2, and its envelope $\widehat{\mathrm{F}}_{p,q}$, along with the deep ties it exhibits with the game via the game's recurrence relations, is elaborated on in §2.3. The results forming the crux of this paper (apart from Theorem 1.1 that has already been stated) are stated in Proposition 2.2, Proposition 2.3 and Proposition 2.4, while the lemma required to prove Proposition 2.2 and Proposition 2.3 is discussed in §3. Mathematically, the most significant section of this paper is §4, in which the method of weight functions is employed to prove Proposition 2.4, and the pivotal steps via which this is accomplished are outlined in §4.1 through §4.7. We go on to provide a formal, game theoretic formulation of our percolation game, along with a detailed description of relevant game theoretic terminology, in §5. In §6, we include discussions on several directions in which future research closely relating to this paper can be carried out, such as examining the probabilities of draw in games where the set of actions is generalized, or where arbitrary mover-sequences are considered, or where the event of draw itself is generalized, percolation games on lattices of higher dimensions, possible monotonicity properties of the probabilities of draw, studying values of percolation games etc. Finally, in §7, we speculate on the various other settings (both in terms of percolation games and 'game theoretic' PCAs that we introduce in §7) in which we believe that the technique of weight functions can be applied to explore the probability of draw for the game under consideration / ergodicity of the PCA under consideration.

2 The principal objects studied in this paper

2.1 Our percolation game

We begin by describing, in detail, the percolation game that we briefly dwelt on in §1.1. The permitted moves are illustrated via the directed (as indicated by the arrowheads), dashed, black lines in Figure 1.

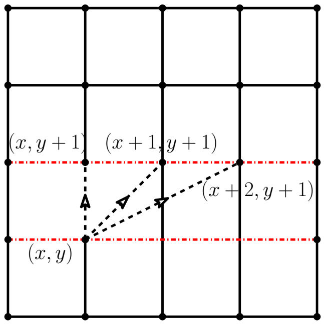

Figure 1: $\mathrm{Out}(x, y)$ in our percolation game.

We recall from §1.1 that each vertex (or site) of $\mathbb{Z}^2$ is assigned, independent of all else, a label that is a trap with probability p, a target with probability q, and open with probability $r = 1 - p - q$, where $(p, q) \in \mathcal{S}$, with $\mathcal{S}$ as in (1.1). Starting from an initial vertex, the two players take turns to move the token from its current position, say (x, y), to any vertex in $\mathrm{Out}(x, y)$. A player wins if she is able to move the token to a target or if her opponent is forced to move it to a trap. We clarify here that, once $\mathbb{Z}^2$ has been endowed with a trap / target / open labeling, the assignment is revealed in its entirety to both the players *before* the game begins (hence, this is a *perfect information game*). Thereafter, to say that the corresponding game is won by a player is to assert that she has a strategy which, when employed, allows her to win no matter what strategy her opponent adopts. The game continues for as long as the token does not land on a site that is marked either a target or a trap, and this could happen indefinitely, leading to a draw. The primary question of interest to us is the same as that in [25], i.e. for what values of the parameters p and q the game exhibits a positive probability of draw. As previously mentioned, the proof pivots upon the connection this game has with the envelope PCA $\widehat{\mathrm{F}}_{p,q}$ (to be introduced formally in §2.3).

In this context, we mention [4] that studies the following two-player combinatorial game on any graph: starting from an initial vertex, the players take turns to move a token, where a move involves relocating the token from the vertex where it is currently situated to a neighbor of that vertex that has not yet been visited. A player who is unable to move loses the game. On $\mathbb{Z}^2$, in which odd and even sites, independently, are marked closed (i.e. forbidden from being visited by the token) with probabilities p and q respectively, it is shown that the game has probability 0 of ending in a draw provided closed sites of one parity are sufficiently rare compared to closed sites of the other parity. This question, however, remains open when the percolation parameters p and q are equal. Motivations for studying the games addressed in [4] include their deep connections with maximum-cardinality matchings in graphs, and in particular, the ways in which draws in these games relate to sensitivity of such matchings to boundary conditions.

2.2 Our PCA

Recall that we have alluded to the specific PCA $F_{p,q}$ that is of interest to us in §1.1, and included a brief description of PCAs in general in §1.2. A PCA F defined on the lattice $\mathbb{Z}^d$ (and hence referred to as a d-dimensional PCA), for some $d \in \mathbb{N}$, comprises a finite set of states $\mathcal{A}$ that is called its *alphabet*, a finite set of indices $\mathcal{N} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m\} \subset \mathbb{Z}^d$ that is called its *neighborhood*, and a stochastic matrix $\varphi : \mathcal{A}^m \times \mathcal{A} \to [0, 1]$ that is called its (random) *local update rule*. Given a *configuration* η in the *state space* $\Omega = \mathcal{A}^{\mathbb{Z}^d}$, we apply F to η, obtaining a (random) configuration $F\eta$, in which the state $F\eta(\mathbf{x})$ of the site

$\mathbf{x} \in \mathbb{Z}^d$ is a random variable whose probability distribution is given by

$$\mathbf{P}[F\eta(\mathbf{x}) = b \big| \eta(\mathbf{x}+\mathbf{y}_i) = a_i \text{ for all } 1 \leqslant i \leqslant m] = \varphi(a_1, a_2, \dots, a_m; b) \text{ for all } b \in \mathcal{A}, \quad (2.1)$$

for any $a_1, a_2, \dots, a_m \in \mathcal{A}$. Here, by definition of stochastic matrices, for all $a_1, a_2, \dots, a_m, b \in \mathcal{A}$, we have $\varphi(a_1, a_2, \dots, a_m; b) \geqslant 0$ and $\sum_{b\in\mathcal{A}} \varphi(a_1, a_2, \dots, a_m; b) = 1$. The updation from $\eta(\mathbf{x})$ to $F\eta(\mathbf{x})$ happens independently over all sites $\mathbf{x} \in \mathbb{Z}^d$. Since we consider discrete-time PCAs, it makes sense to indicate by η_t the configuration at time $t \in \mathbb{N}_0$, so that $\eta_{t+1} = F\eta_t$ for all $t \in \mathbb{N}_0$. We call a PCA *elementary* when it is defined on $\mathbb{Z}$ (i.e. $d = 1$) and $|\mathcal{N}| = |\mathcal{A}| = 2$. We refer the reader to [[34], §2] and [33] for excellent expositions on PCAs in general.

Next, we come to a brief discussion regarding the notion of ergodicity of a d-dimensional PCA. To begin with, we let $\mathcal{F}$ denote the σ-field that is generated by the cylinder sets of $\Omega = \mathcal{A}^{\mathbb{Z}^d}$, and we let $\mathbb{D}$ denote the set of all probability measures supported on Ω and defined with respect to $\mathcal{F}$. We emphasize here that every probability measure on Ω that is henceforth mentioned belongs to the set $\mathbb{D}$. We define $F^t\eta = F(F^{t-1}\eta)$ for $\eta \in \Omega$ and $t \in \mathbb{N}$ (in particular, $F^1\eta = F\eta$). In other words, $F^t\eta$ is the (random) configuration that is obtained by applying F sequentially t times to the initial configuration η. These definitions extend naturally to random η, and if η follows the probability distribution μ (that belongs to $\mathbb{D}$), we let $F^t\mu$ (simply written $F\mu$ when $t = 1$) denote the probability distribution of the (also random) configuration $F^t\eta$.

Definition 2.1. We say that μ is a *stationary* or *invariant* measure for a PCA F if $F\mu = \mu$ (in other words, the pushforward measure induced by F is the same as the original measure). We call the PCA F *ergodic* if it has a *unique* stationary measure, say μ, which is *attractive*, i.e. for *every* probability measure ν on Ω, the sequence $F^t\nu$ converges weakly to μ as $t \to \infty$.

Consider $d = 1$. Given $y \in \mathbb{Z}$ and a configuration η, we define the configuration $T^y\eta$, with $T^y\eta(x) = \eta(x+y)$ for all $x \in \mathbb{Z}$, as the *translation* or *shift* of η by y. We say that a probability measure μ belonging to $\mathbb{D}$ is *translation-invariant* or *shift-invariant* if for every subset B measurable with respect to the σ-field $\mathcal{F}$ introduced earlier, and every $y \in \mathbb{Z}$, we have $\mu(B) = \mu(T^yB)$, where $T^yB = \{T^y\eta : \eta \in B\}$. Given a configuration η, we denote by η^R, with $\eta^R(x) = \eta(-x)$ for all $x \in \mathbb{Z}$, the *reflection* of η. We say that a probability measure μ belonging to $\mathbb{D}$ is *reflection-invariant* if for every subset B measurable with respect to $\mathcal{F}$, we have $\mu(B) = \mu(B^R)$, where $B^R = \{\eta^R : \eta \in B\}$. These notions will be of use to us in the sequel (see §4).

We now come to a detailed description of the PCA $F_{p,q}$, with parameters $(p, q) \in \mathcal{S}$. This is a 1-dimensional PCA, with alphabet $\mathcal{A} = \{0, 1\}$ and neighborhood $\mathcal{N} = \{0, 1, 2\}$, so that $F_{p,q}\eta(n)$ is a random variable whose probability distribution is a function of $\eta(n)$, $\eta(n+1)$ and $\eta(n+2)$ for all $n \in \mathbb{Z}$. More precisely, slighty tweaking the notation introduced in (2.1), the stochastic matrix $\varphi_{p,q} : \mathcal{A}^3 \times \mathcal{A} \to [0, 1]$ for this PCA is defined via the equations:

$$\varphi_{p,q}(0,0,0;b) = \begin{cases} p & \text{if } b = 0, \\ 1-p & \text{if } b = 1, \end{cases} \quad (2.2)$$

and

$$\varphi_{p,q}(a_0, a_1, a_2; b) = \begin{cases} 1-q & \text{if } b = 0, \\ q & \text{if } b = 1, \end{cases} \quad \text{for all } (a_0, a_1, a_2) \in \mathcal{A}^3 \setminus \{(0,0,0)\}. \quad (2.3)$$

The (stochastic) update rule for the automaton $F_{p,q}$ can be illustrated pictorially via Figure 2.

As defined in [34], given $\epsilon > 0$, a PCA, say Φ, is said to be an ϵ-*perturbation* of a deterministic CA, say F, if Φ and F share the same alphabet, say S, and the same neighborhood, and the stochastic update rule φ of Φ satisfies $\varphi(a_1, a_2, \ldots, a_m; f(a_1, a_2, \ldots, a_m)) \geqslant 1 - \epsilon$ for all $(a_1, a_2, \ldots, a_m) \in S^m$, where f is the local update rule for F. In other words, given any input $(a_1, a_2, \ldots, a_m) \in S^m$, the PCA Φ outputs the same symbol as F with probability at least $1 - \epsilon$, and alters it with the remaining probability. The PCA $F_{p,q}$ described above can be derived from a deterministic CA, F, and a stochastic noise, θ, via such a perturbation. The local update rule f of F is given by $f(a_0, a_1, a_2) = 1 - \max\{a_0, a_1, a_2\}$, for all $a_0, a_1, a_2 \in \mathcal{A}$, while the noise $\theta(a, b)$, for $a, b \in \{0, 1\}^2$, transforms 1 into 0 with probability p (i.e. $\theta(1, 0) = p$) and transforms 0 into 1 with probability q (i.e. $\theta(0, 1) = q$). It is easily verified that $\varphi_{p,q}(a_0, a_1, a_2; b) = \theta(f(a_0, a_1, a_2), b)$ for all $a_0, a_1, a_2, b \in \mathcal{A}$. Moreover, we see that $\theta(a, b) = (1 - \epsilon)\delta_a(b) + \epsilon g(b)$ for $a, b \in \mathcal{A}$, where $\epsilon = p + q$, $\delta_a(b) = 1$ when $b = a$ and $\delta_a(b) = 0$ when $b \neq a$, and $g(0) = 1 - g(1) = \frac{p}{p+q}$. Thus, our PCA $F_{p,q}$ is obtained from the CA F via perturbation using the memoryless zero-range noise θ (we refer the reader to [34] for further reading on such noise-perturbed CAs).

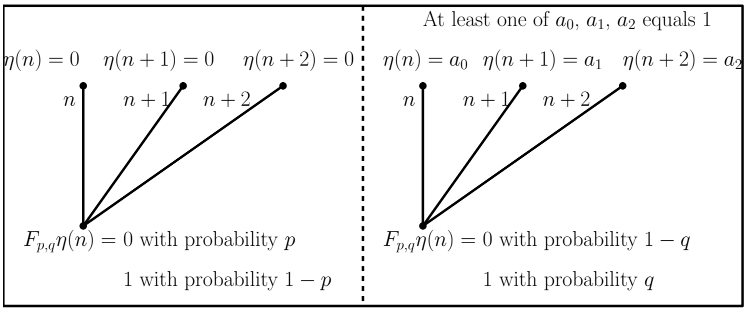

Figure 2: The stochastic rules that define our PCA $F_{p,q}$.

The PCA $A_{p,q}$ that is studied in [25] is an elementary one that bears some resemblance with $F_{p,q}$ in that, following the notation in (2.1), its stochastic matrix $\varphi : \mathcal{A}^2 \times \mathcal{A}$ is defined by the equations:

$$\varphi(0,0;b) = \begin{cases} p & \text{if } b = 0, \\ 1-p & \text{if } b = 1, \end{cases} \quad \text{and}$$

$$\varphi(a_0,a_1;b) = \begin{cases} 1-q & \text{if } b = 0, \\ q & \text{if } b = 1, \end{cases} \quad \text{for all } (a_0,a_1) \in \mathcal{A}^2 \setminus \{(0,0)\}.$$

We draw the reader's attention to the primary contrast between $F_{p,q}$ and $A_{p,q}$: whereas in $A_{p,q}$, we draw upon information regarding the states of 2 consecutive sites, n and $n+1$, in order to decide the (random) updated state of the site n, in $F_{p,q}$, we draw upon information regarding the states of 3 consecutive sites, n, $n+1$ and $n+2$, to decide the (random) updated state of the site n. It is, therefore, expected that the latter will have a somewhat-more-involved underlying dependence structure, and establishing ergodicity results for $F_{p,q}$ indeed proves to be a considerably more challenging feat compared to that for $A_{p,q}$, at least as far as using the method of weight functions proposed and implemented in [25] is concerned.

Some discussion on how important the question of ergodicity (or the lack thereof) of PCAs is, and how difficult this question is to resolve under various circumstances, is now in order. It has been found to be notoriously difficult to construct a CA whose trajectories, starting from different initial conditions and evolving under repeated applications of the local update rule, remain distinguishable from each other if even the slightest positive noise is incorporated into the CA (for instance, see [34]). In other words, most CAs tend to forget their initial conditions under the influence of even the smallest amount of noise, and this is reflected in the ergodicity of the resulting PCA. The renowned *positive rates conjecture* states that *all* PCAs defined on $\mathbb{Z}$ and satisfying $\varphi(a_1, a_2, \ldots, a_{|\mathcal{N}|}; b) > 0$ for all $a_1, a_2, \ldots, a_{|\mathcal{N}|}, b \in \mathcal{A}$ (referred to as the *positive rates* condition) are ergodic. An extremely complicated example proposed by [17] refutes this conjecture, but the fascinating question still remains as to whether all sufficiently simple, naturally occurring 1-dimensional PCAs with positive rates are ergodic. There are, however, examples of d-dimensional PCAs for $d \geqslant 2$, such as Glauber dynamics for the Ising model at low temperatures, that are known to be non-ergodic. We note here that both $F_{p,q}$ and $A_{p,q}$ have positive rates as long as *both* p and q are strictly positive.

Multiple sources (see discussions in [11], [25] and [9]) reiterate that in general, even if the answer can be guessed from heuristics or simulations, *rigorously* proving whether a given PCA is ergodic or not is a notoriously difficult problem, and is shown to be algorithmically undecidable in [10] and [13]. Under the assumption of left-right symmetry (which guarantees $\varphi(1, 0; 0) = \varphi(0, 1; 0)$), an elementary PCA is determined by the parameters $\varphi(0, 0; 0)$, $\varphi(1, 1; 0)$ and $\varphi(1, 0; 0)$ (recall these notations from (2.1)). The many existing techniques that have been developed to study ergodicity can take care of ergodicity questions for such PCAs over more than 90% of the volume of the cube $[0, 1]^3$ defined by these 3 parameters ([13]). However, when p and q are small, $A_{p,q}$ belongs to a domain of this cube where none of these techniques works, and this is where [25] comes in with their brilliant idea of employing weight functions. To the best of our knowledge, the problem of establishing ergodicity results for $F_{p,q}$, for *all* $(p, q) \in \mathcal{S}$, is an even more challenging one that has, so far, remained open, and we, in this paper, utilize the method of weight functions to provide a concrete proof of the first assertion of Theorem 1.1.

Apart from being an interesting object to study in its own right, motivation to investigate $F_{p,q}$ can be found from the connections that such PCAs have with the problem of enumerating *directed animals*, as was first pointed out by [12], and further explored in [6]. The relation between PCAs and the problem of enumerating directed animals is elucidated upon in [33]. In [8], the generating functions for directed animals on square and triangular lattices, as well as on decorated square and triangular lattices, are obtained, further demonstrating the usefulness of these PCAs.

In [6], a directed graph G is considered, in which a vertex v is called a *successor* to a vertex u if there exists a directed edge *from* u *to* v. Given any subset S of vertices of G, a *directed animal* A with *source* S is a subset of vertices of G such that $S \subset A$, and every $v \in A \setminus S$ can be reached from S via a directed path all of whose vertices are in A. The number of vertices in A, denoted $|A|$, is referred to as the *area* of A. In order to count the number of possible directed animals on G with a given source S and a given area n, a suitable *generating function* $G_S^G(x) = x^{|S|} + \cdots$ is considered, where $\cdots$ represents a sum of monomials whose degrees are at least $|S| + 1$.

Next, [6] defines a *particle system* or *gas occupation* on G as a map X from the set of vertices of G to $\{0, 1\}$, with respect to which any vertex u is said to be *occupied* if $X_u = 1$. When X is random, this is referred to as a *random gas model*. A random colouring of the vertices of G with colours a and b is considered. Letting C_u denote the colour of the vertex u for each u in G, the random variables C_u are independent and identically distributed, $C_u = a$ with probability p and $C_u = b$ with probability $1 - p$. The desired gas model is now defined with respect to this random colouring, as follows: for each vertex u of G,

$$X_u = \begin{cases} 0 & \text{if } C_u = a, \\ \prod_{v \text{ successor of } u} (1 - X_v) & \text{if } C_u = b. \end{cases} \tag{2.4}$$

One of the main results of [6] states that if R_u, for any vertex u of G, is the radius of convergence of the generating function $G^G_{\{u\}}(x)$, and $(1-p) \in [0, R_u)$, then $\mathbf{E}[X_u] = -G^G_{\{u\}}(-1+p)$. Therefore, knowledge about the random colouring scheme described above is valuable for the understanding of these generating functions.

It is fairly immediate that the above-mentioned random colouring scheme can be represented by a special case of our PCA $F_{p,q}$ when we consider $q = 0$, and in place of G, we consider either a 2-dimensional infinite directed lattice in which the successors of any site (x, y) are $(x, y+1)$, $(x+1, y+1)$ and $(x+2, y+1)$ (see Figure 3), or a 2-dimensional infinite directed lattice in which the successors of any site (x, y) are $(x+2, y)$, $(x+1, y+1)$ and $(x, y+2)$ (see Figure 4).

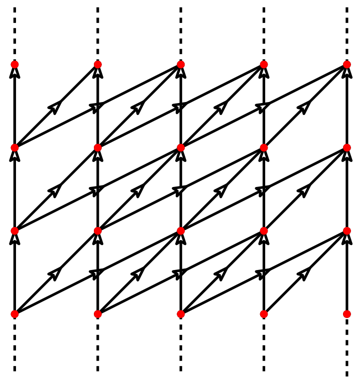

Figure 3:

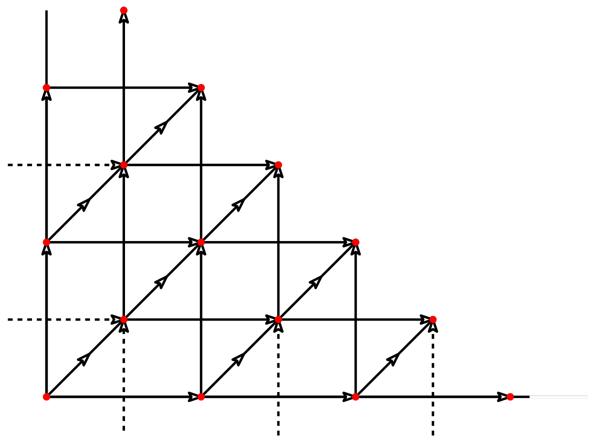

Figure 4:

Referring to [9], we know that a *binary symmetric channel*, denoted BSC_δ where $0 < \delta < 1$ is a pre-specified parameter known as the *crossover probability*, takes as its input a bit and flips it with probability δ (and keeps it unchanged with probability $1-\delta$). For the case where $p = q = \delta$, we see that $F_{p,q}$ can be represented as

$$F_{\delta,\delta}\eta(n) = \mathrm{BSC}_\delta \left(\mathrm{NAND}\left(\eta(n), \eta(n+1), \eta(n+2)\right)\right), \tag{2.5}$$

where the NAND function outputs the value 1 when all three inputs are 0, and outputs the value 0 when at least one of the three inputs equals 1. As outlined in § 3.2 of [9], the problem of *broadcasting of information on* 2*-dimensional grids* showcases another usefulness of such PCAs. Let us consider the problem of broadcasting information on the two 2-dimensional infinite directed grid shown in Figure 5. The origin O has one bit of information, and we are interested in broadcasting this bit to the entire infinite directed lattice. Let us associate with each vertex of the grid the index pair (t, i), with $t \geqslant 0$ and $0 \leqslant i \leqslant 2t$. At any time $t > 0$, each vertex with index (t, i), for $2 \leqslant i \leqslant 2t-2$, receives a bit from each of the vertices indexed $(t-1, i-2)$, $(t-1, i-1)$ and $(t-1, i)$, applies the NAND function to these bits, and then implements BSC_δ. The vertex indexed $(t, 1)$ receives bits from the vertices $(t-1, 0)$ and $(t-1, 1)$, the vertex indexed $(t, 2t-1)$ receives bits from the vertices $(t-1, 2t-3)$ and $(t-1, 2t-2)$, the vertex indexed $(t, 0)$ receives a bit from the vertex $(t-1, 0)$, and the vertex indexed $(t, 2t)$ receives a bit from the vertex $(t-1, 2t-2)$. From (2.5), it is evident that if we view the coordinate t as time, understanding this broadcasting problem is similar to studying the PCA $F_{\delta,\delta}$, except that broadcasting on such a grid involves bounded-length configurations and its behaviour is different at the boundary.

As a third motivation for studying the PCA $F_{p,q}$, we point out, as discussed right after (2.3), that $F_{p,q}$ can be obtained by perturbing a deterministic CA by a memoryless

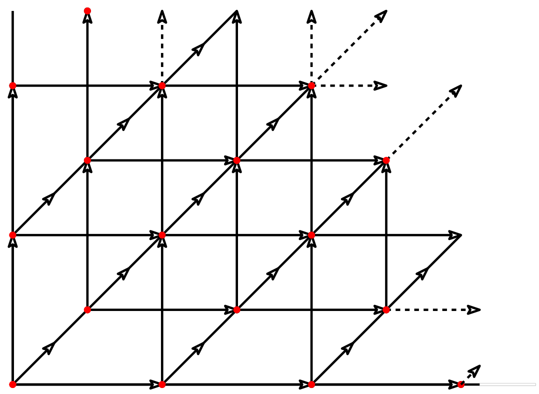

Figure 5: Broadcasting of information on this grid resembles the PCA $F_{p,q}$ with $p = q = \delta$.

zero-range noise (once again, we refer the reader to [34]). The importance of noisy CAs, as a subclass of PCAs, lies in their use in studying the reliability of computations against the presence of noise. The study of low-noise PCAs presents us with the same kind of challenges that we face when studying low-temperature models of statistical mechanics. In particular, studying the ergodicity, or lack thereof, of such PCAs, is intimately tied to the phenomenon of phase transition at low temperatures. This, along with the previous discussions, serve to motivate our investigation of the PCA $F_{p,q}$.

2.3 Our envelope PCA and its relation to our percolation game

We begin by deducing certain recurrence relations that arise naturally in the percolation game described in §2.1. Once $\mathbb{Z}^2$ has been endowed with the trap / target / open labeling, we define a site (x, y) to be in the class W if the game that begins with (x, y) as the initial vertex is won by the player who plays the first round. We define (x, y) to be in the class L if the game that begins with (x, y) as the initial vertex is lost by the player who plays the first round, and we define (x, y) to be in the class D if the game that begins with (x, y) as the initial vertex results in a draw. In particular, if (x, y) is a trap, then we place it in W, and if it is a target, we place it in L. The intuition behind these conventions is as follows: one may imagine an "unseen" round that takes place *before* the actual game begins, in which the player who is supposed to play the second round of the actual game moves the token from somewhere else to (x, y). Thus, if (x, y) is a trap (respectively a target), she loses (respectively wins) even before the game begins, implying that the player who plays the first round of the actual game wins (respectively loses).

For every $k \in \mathbb{Z}$, we denote by $H_k = \{(x, k) : x \in \mathbb{Z}\}$ the horizontal line containing all sites whose y-coordinate equals k. These lines have been illustrated in red in Figure 6. From the moves permitted in our game, it follows that for any $k \in \mathbb{Z}$, if all sites (x, y) that

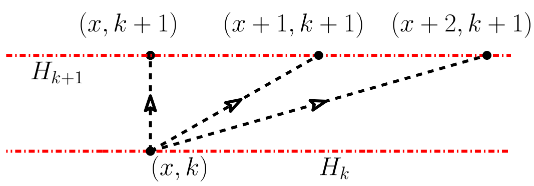

Figure 6: Illustrating deduction of the recurrence relations.

lie on the horizontal line H_{k+1} have already been categorized into the classes W, L and

D, then this information, along with the pre-assigned labels of trap / target / open to the sites on H_k, is enough to determine the classes to which the sites lying on H_k belong.

Using Figure 6 as a reference, we draw the following conclusions, assuming that (x, y) is our initial vertex:

1. If each vertex of $\mathrm{Out}(x, y)$ belongs to W, then no matter which of these vertices the first player moves the token to from (x, y), the second player wins. We thus have the following two possibilities:
   (a) either the vertex (x, y) has been marked a trap and hence belongs to W, which happens with probability p,
   (b) or else the game that begins from (x, y) results in a loss for the first player, so that (x, y) is classified into L with the remaining probability $1 - p$.
2. If at least one of the vertices of $\mathrm{Out}(x, y)$ belongs to L, the first player moves the token from (x, y) to *this* vertex, making the second player lose. We thus have the following two possibilities:
   (a) either (x, y) has been marked a target and hence belongs to L, which happens with probability q,
   (b) or else the game that begins from (x, y) results in a win for the first player, so that (x, y) is classified into W with the remaining probability $1 - q$.
3. The final scenario is where *none* of the vertices of $\mathrm{Out}(x, y)$ belongs to L but at least one of them belongs to D. In this case, we have the following three possibilities:
   (a) either (x, y) has been marked a trap and hence belongs to W, which happens with probability p,
   (b) or (x, y) has been marked a target and hence belongs to L, which happens with probability q,
   (c) or else the game that begins from (x, y) results in a draw (since the first player moves the token from (x, y) to the vertex in $\mathrm{Out}(x, y)$ which is in D), thus placing (x, y) in the class D with the remaining probability $r = 1 - p - q$.

We further note that, conditioned on the classification of the vertices that lie on H_{k+1} into the classes W, L and D, the (random) class which a vertex lying on H_k gets sorted into via the above-mentioned rules is independent of all other vertices on H_k.

The above recurrence relations are the key to establishing a connection between our game and the envelope PCA $\widehat{\mathrm{F}}_{p,q}$ that we are now ready to define. Let us identify W with the symbol 0, L with the symbol 1 and D with the symbol $?$ (in other words, we label a vertex 0 if it has been classified into W, 1 if it has been classified into L, and $?$ if it has been classified into D). For any $k \in \mathbb{Z}$, we identify H_k with $\mathbb{Z}$ by identifying (x, k) on H_k with x on $\mathbb{Z}$ for each $x \in \mathbb{Z}$. This allows us to represent the recurrence relations listed above via a PCA $\widehat{\mathrm{F}}_{p,q}$ that is endowed with the alphabet $\hat{\mathcal{A}} = \{0, 1, ?\}$, the neighborhood $\mathcal{N} = \{0, 1, 2\}$, and the stochastic matrix $\widehat{\varphi}_{p,q} : \hat{\mathcal{A}}^3 \times \hat{\mathcal{A}} \to [0, 1]$ defined via the equations:

$$\widehat{\varphi}_{p,q}(0,0,0;b) = \begin{cases} p & \text{if } b = 0, \\ 1-p & \text{if } b = 1, \end{cases} \tag{2.6}$$

$$\widehat{\varphi}_{p,q}(a_0,a_1,a_2;b) = \begin{cases} 1-q & \text{if } b = 0, \\ q & \text{if } b = 1, \end{cases} \quad \text{for all } (a_0,a_1,a_2) \in \hat{\mathcal{A}}^3 \setminus \{0,?\}^3 \tag{2.7}$$

and

$$\widehat{\varphi}_{p,q}(a_0,a_1,a_2;b) = \begin{cases} p & \text{if } b = 0, \\ q & \text{if } b = 1, \\ r = 1-p-q & \text{if } b = ?, \end{cases} \quad \text{for all } (a_0,a_1,a_2) \in \{0,?\}^3 \setminus (0,0,0). \tag{2.8}$$

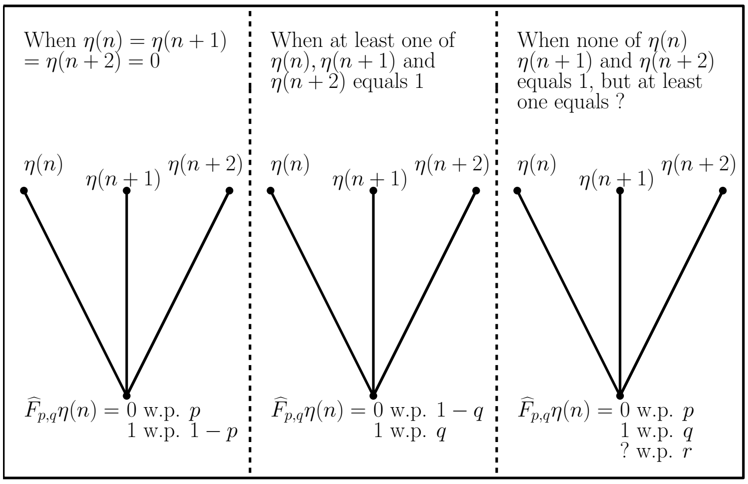

Figure 7: The stochastic rules that define our envelope PCA $\widehat{\mathrm{F}}_{p,q}$.

We illustrate $\widehat{\varphi}_{p,q}$ in Figure 7. To clarify further, the classification of the vertices on H_{k+1} into W, L and D yields a configuration $\eta \in \hat{\mathcal{A}}^{\mathbb{Z}}$ via the identifications described above, and the classification of the vertices on H_k into W, L and D via the game's recurrence relations can then be represented by $\widehat{\mathrm{F}}_{p,q}\, \eta$.

If we endow $\hat{\mathcal{A}}$ with the total order $0 \prec ? \prec 1$ and define $1-? = ?$, then (similar to the way we interpret $F_{p,q}$ in §2.2) the envelope PCA $\widehat{\mathrm{F}}_{p,q}$ can be derived, via random perturbations, from a deterministic CA, $\widehat{\mathrm{F}}$, and a random noise $\widehat{\theta}$, as follows. The local update rule $\widehat{\mathrm{f}}$ of $\widehat{\mathrm{F}}$ is given by $\widehat{\mathrm{f}}(a_0, a_1, a_2) = 1 - \max\{a_0, a_1, a_2\}$, whereas the noise $\widehat{\theta}(a, b)$, for $a, b \in \hat{\mathcal{A}}$, transforms 0 to 1 with probability q (i.e. $\widehat{\theta}(0, 1) = q$) and keeps it unchanged with probability $1-q$, transforms 1 to 0 with probabililty p (i.e. $\widehat{\theta}(1, 0) = p$) and keeps it unchanged with probability $1-p$, and transforms ? to 1 with probability q (i.e. $\widehat{\theta}(?, 1) = q$), ? to 0 with probability p (i.e. $\widehat{\theta}(?, 0) = p$), and keeps ? unchanged with probability $r = 1-p-q$. It is now straightforward to check that $\widehat{\mathrm{F}}_{p,q}$ is obtained from $\widehat{\mathrm{F}}$ by injecting into it the memoryless zero-range noise $\widehat{\theta}$ (again, we refer the reader to [34] for a detailed reading on PCAs obtained via perturbations of deterministic CAs by random noises).

Although we have argued in §2.3 how the recurrence relations deduced from our percolation game give rise to $\widehat{\mathrm{F}}_{p,q}$, there is, in fact, another angle from which one may motivate the emergence of $\widehat{\mathrm{F}}_{p,q}$ (we mention here that the term "envelope" was introduced in [10]) – namely, the use of $\widehat{\mathrm{F}}_{p,q}$ in coupling two (random) configurations obtained via (possibly repeated) applications of the PCA $F_{p,q}$, starting from two different initial configurations. The symbol ? is utilized to populate sites whose actual values may differ between these two coupled configurations.

We dwell here a little longer on the notion of envelope PCAs. In [10], a random algorithm known as a *perfect sampling procedure* is proposed which, when repeatedly applied, can estimate the (unique) stationary probability distribution corresponding to an ergodic PCA F with arbitrary precision. When the universe of F is finite (e.g. when $\mathbb{Z}/n\mathbb{Z}$ constitute the cells for F), a new PCA $\widehat{\mathrm{F}}$ on an extended alphabet, referred to as an *envelope PCA to F*, is introduced, which is then run on a single initial configuration to obtain a perfect sampling procedure for F. When the universe for F is infinite (e.g. when $\mathbb{Z}^d$ constitute the cells for F, making the universe uncountably infinite), [10] shows that an efficient perfect sampling procedure can be obtained when $\widehat{\mathrm{F}}$ is ergodic.

When the alphabet for F is $\mathcal{A} = \{0, 1\}$, the alphabet for $\widehat{\mathrm{F}}$ is $\hat{\mathcal{A}} = \{0, 1, ?\}$. The universe

E and neighborhood $\mathcal{N}$ for $\widehat{\mathrm{F}}$ remain the same as those for F. Given a configuration $\eta = (\eta(x) : x \in E)$ in $\hat{\mathcal{A}}^E$, one thinks of η as a configuration in $\mathcal{A}^E$ as follows: if $x \in E$ is a cell such that it is not known which symbol from $\mathcal{A}$ occupies it, we imagine that it is occupied by the symbol $?$. It is to be noted that when $\eta \in \mathcal{A}^E$, i.e. η is a configuration which is devoid of the symbol $?$, the envelope PCA $\widehat{\mathrm{F}}$ acts on it in exactly the same way as F acts on it.

It is now time to connect the principal components of this paper, namely the game, the PCA $F_{p,q}$, and its envelope $\widehat{\mathrm{F}}_{p,q}$, with one another. This connection is established via the following three results.

Proposition 2.2. For every $(p, q) \in \mathcal{S}$, the percolation game with parameters p and q has probability 0 of culminating in a draw if and only if the PCA $F_{p,q}$ is ergodic.

Proposition 2.3. The PCA $F_{p,q}$ is ergodic if and only if the corresponding envelope PCA $\widehat{\mathrm{F}}_{p,q}$ is ergodic.

Proposition 2.4. For each $(p, q) \in \mathcal{S}$, the envelope PCA $\widehat{\mathrm{F}}_{p,q}$ admits no stationary distribution μ that assigns positive probability to the symbol $?$. To put it formally, the probability of the event $\eta(x) =?$, where η is a random configuration with law μ, is 0 for every $x \in \mathbb{Z}$.

We implement the first assertion of Lemma 3.1 (that concerns itself with a couple of results regarding stochastic domination, as stated in §3 below) and the argument used in proving Proposition 2.1 of [25] to establish Proposition 2.3. Next, an argument identical to that used in proving Proposition 2.2 of [25] yields a proof of Proposition 2.2. We are now left with the task of proving Proposition 2.4.

3 An important lemma before we embark on a proof of Proposition 2.4

Recall from §2.3 that $\hat{\mathcal{A}} = \{0, ?, 1\}$, and borrowing from the definitions in §2.2, we let $\mathbb{D}$ denote the set of all probability measures on $\Omega = \hat{\mathcal{A}}^{\mathbb{Z}}$ that are defined with respect to the σ-field $\mathcal{F}$ generated by the cylinder sets of Ω.

Lemma 3.1. *Let μ and $\widetilde{\mu}$ be two probability distributions in $\mathbb{D}$. Let $\preceq$ denote the stochastic domination (on $\mathbb{D}$) with respect to the coordinate-wise total order induced by the ordering $0 \prec ? \prec 1$, and let $\mu \preceq \widetilde{\mu}$. Then $\widehat{\mathrm{F}}_{p,q}\,\widetilde{\mu} \preceq \widehat{\mathrm{F}}_{p,q}\,\mu$.*

Let μ and $\widetilde{\mu}$ be two probability distributions in $\mathbb{D}$. Let $\trianglelefteq$ denote the stochastic domination (on $\mathbb{D}$) with respect to the coordinate-wise partial order induced by $0 \lhd ? \rhd 1$, and let $\mu \trianglelefteq \widetilde{\mu}$. Then $\widehat{\mathrm{F}}_{p,q}\,\mu \trianglelefteq \widehat{\mathrm{F}}_{p,q}\,\widetilde{\mu}$.

Proof. We begin with (deterministic) configurations η and $\widetilde{\eta}$ in $\hat{\mathcal{A}}^{\mathbb{Z}}$ such that $\eta(x) \preceq \widetilde{\eta}(x)$ for all $x \in \mathbb{Z}$. We consider two copies of $\mathbb{Z}^2$, in both of which we fix the *same* assignment, say ω, of trap / target / open labels to the vertices lying on H_k. In the first copy of $\mathbb{Z}^2$, we assign the state $\eta(x)$ to the site $(x, k+1)$ (that belongs to H_{k+1}), for each $x \in \mathbb{Z}$, whereas in the second copy of $\mathbb{Z}^2$, we assign the state $\widetilde{\eta}(x)$ to the site $(x, k+1)$, for each $x \in \mathbb{Z}$.

The state, say $\tau(x)$, assigned to the site (x, k) (that belongs to H_k), for each $x \in \mathbb{Z}$, in the first copy of $\mathbb{Z}^2$, is deduced using the states $\eta(x)$, $\eta(x+1)$ and $\eta(x+2)$, and the label $\omega(x, k)$, according to the following rules:

1. if $\omega(x, k)$ reads trap, then $\tau(x) = 0$,
2. if $\omega(x, k)$ reads target, then $\tau(x) = 1$,
3. if $\omega(x, k)$ reads open, and
   (a) $\eta(x) = \eta(x+1) = \eta(x+2) = 0$, then $\tau(x) = 1$,

(b) at least one of $\eta(x)$, $\eta(x+1)$ and $\eta(x+2)$ equals 1, then $\tau(x) = 0$,
(c) none of $\eta(x)$, $\eta(x+1)$ and $\eta(x+2)$ equals 1, but at least one of them equals ?, then $\tau(x) =?$.

The state, say $\widetilde{\tau}(x)$, assigned to (x, k), for each $x \in \mathbb{Z}$, in the second copy of $\mathbb{Z}^2$, is deduced using the states $\widetilde{\eta}(x)$, $\widetilde{\eta}(x+1)$ and $\widetilde{\eta}(x+2)$, and the label $\omega(x, k)$, following rules analogous to the above.

We now compare $\tau(x)$ and $\widetilde{\tau}(x)$ for each $x \in \mathbb{Z}$. Since the same assignment, ω, of trap / target / open labels to the vertices of H_k is used in each copy of $\mathbb{Z}^2$, we need only carry out this comparison when $\omega(x, k)$ is open. It suffices to consider $x = 0$.

1. Suppose $\widetilde{\eta}(0) = \widetilde{\eta}(1) = \widetilde{\eta}(2) = 0$, which forces $\eta(0) = \eta(1) = \eta(2) = 0$. In this case, $\tau(0) = \widetilde{\tau}(0) = 1$.
2. Suppose $(\widetilde{\eta}(0), \widetilde{\eta}(1), \widetilde{\eta}(2)) \in \{0, ?\}^3 \setminus \{(0, 0, 0)\}$. In this case, we either have $(\eta(0), \eta(1), \eta(2)) \in \{0, ?\}^3 \setminus \{(0, 0, 0)\}$ or we have $(\eta(0), \eta(1), \eta(2)) = (0, 0, 0)$. In the former situation, $\tau(0) = \widetilde{\tau}(0) =?$, whereas in the latter situation, $\widetilde{\tau}(0) =?$ and $\tau(0) = 1$.
3. Finally, suppose $(\widetilde{\eta}(0), \widetilde{\eta}(1), \widetilde{\eta}(2)) \in \{0, ?, 1\}^3 \setminus \{0, ?\}^3$. The first possibility is that $(\eta(0), \eta(1), \eta(2)) \in \{0, ?, 1\}^3 \setminus \{0, ?\}^3$, in which case $\tau(0) = \widetilde{\tau}(0) = 0$. The second possibility is where $(\eta(0), \eta(1), \eta(2)) \in \{0, ?\}^3 \setminus \{(0, 0, 0)\}$, in which case, $\widetilde{\tau}(0) = 0$ whereas $\tau(0) =?$. The third and final possibility is where $(\eta(0), \eta(1), \eta(2)) = (0, 0, 0)$, in which case $\widetilde{\tau}(0) = 0$ and $\tau(0) = 1$.

These observations let us conclude that $\tau \succeq \widetilde{\tau}$. Given $\mu \preceq \widetilde{\mu}$, we let η and $\widetilde{\eta}$ denote coupled random configurations (defined on the same sample space) such that η follows μ, $\widetilde{\eta}$ follows $\widetilde{\mu}$ and $\eta \preceq \widetilde{\eta}$ almost surely. The above deduction applied to η and $\widetilde{\eta}$ completes the proof of the first assertion of Lemma 3.1.

To prove the second assertion of Lemma 3.1, we consider a very similar set-up as above, with the only distinction being that $\eta(x) \trianglelefteq \widetilde{\eta}(x)$ for all $x \in \mathbb{Z}$. Once again, we compare $\tau(x)$ and $\widetilde{\tau}(x)$ for each $x \in \mathbb{Z}$, and it suffices to consider $x = 0$ and $\omega(0, k)$ open:

1. Suppose $\eta(0) = \eta(1) = \eta(2) = 0$ and $(\widetilde{\eta}(0), \widetilde{\eta}(1), \widetilde{\eta}(2)) \in \{0, ?\}^3 \setminus \{(0, 0, 0)\}$. Then $\tau(0) = 1$ whereas $\widetilde{\tau}(0) =?$.
2. When $(\eta(0), \eta(1), \eta(2)) \in \{0, ?, 1\}^3 \setminus \{0, ?\}^3$, we either have $(\widetilde{\eta}(0), \widetilde{\eta}(1), \widetilde{\eta}(2)) \in \{0, ?\}^3 \setminus \{(0, 0, 0)\}$ or $(\widetilde{\eta}(0), \widetilde{\eta}(1), \widetilde{\eta}(2)) \in \{0, ?, 1\}^3 \setminus \{0, ?\}^3$. In the latter scenario, $\tau(0) = \widetilde{\tau}(0) = 0$, while in the former, we have $\tau(0) = 0$ and $\widetilde{\tau}(0) =?$.

These observations imply $\tau \trianglelefteq \widetilde{\tau}$, which, in turn, yields the conclusion in the second part of Lemma 3.1. □

4 The method of weight functions and the proof of Proposition 2.4

As discussed in §2.2, it is a non-trivial task to establish, rigorously, the ergodicity of PCAs in general. It has also been explicitly stated in [25] that coming up with a suitable weight function or potential function that serves our purpose of proving ergodicity is not an easy feat either.

Before we proceed further, we recall here the definitions of translation-invariant and reflection-invariant probability measures as discussed right after Definition 2.1. Given any 1-dimensional PCA with alphabet $\mathcal{A}$, a finite index set $S = \{y_1, y_2, \dots, y_n\} \subset \mathbb{Z}$, and symbols $a_1, a_2, \dots, a_n$ that belong to $\mathcal{A}$, we call $(a_1 a_2 \dots a_n)_S = \{\eta \in \mathcal{A}^{\mathbb{Z}} : \eta(y_i) = a_i \text{ for all } 1 \leqslant i \leqslant n\}$ a *cylinder set indexed by* S. When a probability measure μ is translation-invariant, and $y_i = k + i$ for all $1 \leqslant i \leqslant n$ and for some $k \in \mathbb{Z}$, we denote the measure $\mu((a_1 a_2 \dots a_n)_S)$ of the cylinder set $(a_1 a_2 \dots a_n)_S$ by simply $\mu(a_1 a_2 \dots a_n)$, since the ‘location’ of S (i.e. the value of k) ceases to be relevant. For instance, if η is a random configuration following the law μ, then for *any* $x \in \mathbb{Z}$, we can let $\mu(?)$ indicate the probability of the event $\eta(x) =?$. Likewise, for *any* $x \in \mathbb{Z}$, we can let $\mu(0?)$ indicate the probability of the event that $\eta(x) = 0$ and $\eta(x+1) =?$, and so on. When μ is both translation-invariant and reflection-invariant, $\mu(a_1 a_2 \dots a_n) = \mu(a_n a_{n-1} \dots a_1)$ for all $a_1, \dots, a_n \in \mathcal{A}$. For instance, $\mu(10?) = \mu(?01)$, since the former is the probability of the event $\eta(-1) = 1, \eta(0) = 0, \eta(1) =?$ whereas the latter is the probability of the event $\eta(-1) =?, \eta(0) = 0, \eta(1) = 1$.

Lemma 4.1. *To prove Proposition 2.4, it suffices to show that under no translation-invariant and reflection-invariant stationary distribution for $\widehat{\mathrm{F}}_{p,q}$ can the symbol $?$ appear anywhere with positive probability, for each $(p,q) \in \mathcal{S}$ (where $\mathcal{S}$ is as defined in (1.1)).*

Proof. Suppose there exists a stationary distribution μ for $\widehat{\mathrm{F}}_{p,q}$ with $\mu(?) > 0$. To prove Lemma 4.1, it now suffices to show the existence of a translation-invariant and reflection-invariant stationary distribution for $\widehat{\mathrm{F}}_{p,q}$ under which the probability of occurrence of the symbol $?$ is strictly positive.

Let $\delta_?$ indicate the configuration that assigns the state $?$ to each $x \in \mathbb{Z}$. By the second assertion of Lemma 3.1, the sequence $\{\widehat{\mathrm{F}}_{p,q}^n \delta_?(?)\}_{n \in \mathbb{N}_0}$ is non-increasing, and $\widehat{\mathrm{F}}_{p,q}^n \delta_?(?) \geqslant \mu(?) > 0$ for each $n \in \mathbb{N}_0$. Thus, any limit point of the Césaro sums of the sequence $\{\widehat{\mathrm{F}}_{p,q}^n \delta_?\}_{n \in \mathbb{N}_0}$ is a translation-invariant and reflection-invariant stationary distribution for $\widehat{\mathrm{F}}_{p,q}$ that assigns a strictly positive probabililty to the occurrence of $?$ (because of our assumption that $\mu(?) > 0$), yielding the desired conclusion. ☐

In the rest of this paper, we consider only μ that belongs to $\mathbb{D}$ (recall from §3) and that is both translation-invariant and reflection-invariant.

We now come to the actual construction of the weight function, which is accomplished via several steps. To begin with, for the sake of brevity, we let $\widehat{**}$ denote the set $\{0,?\}^2 \setminus \{(0,0)\}$ and $\widehat{***}$ the set $\{0,?\}^3 \setminus \{(0,0,0)\}$. We may think of $(\widehat{**})$ as representing the cylinder set in which $(\eta(0), \eta(1)) \in \widehat{**}$, and $(\widehat{***})$ as representing the cylinder set in which $(\eta(0), \eta(1), \eta(2)) \in \widehat{***}$. We write $(S_0 S_1 \dots S_k \widehat{***} S'_0 S'_1 \dots S'_{k'})$ (likewise, $(S_1 S_2 \dots S_k \widehat{**} S'_1 S'_2 \dots S'_{k'})$) to indicate the cylinder set in which $\eta(i) \in S_i$ for all $0 \leqslant i \leqslant k$, $(\eta(k+1), \eta(k+2), \eta(k+3)) \in \widehat{***}$, and $\eta(k+4+i) \in S'_i$ for all $0 \leqslant i \leqslant k'$, for any subsets $S_0, \dots, S_k, S'_0, \dots, S'_{k'}$ of $\{0,?,1\}$. When $S_i = \{a_i\}$ is a singleton for some $0 \leqslant i \leqslant k$, we replace S_i in the above notation by simply a_i (and likewise when S'_i is a singleton for some $0 \leqslant i \leqslant k'$).

Keeping the reader’s convenience in mind and before we plunge into the intricate technicalities of the weight function derivation, we briefly dwell here on how we plan to accomplish the task at hand, i.e. proving Proposition 2.4, using the method of weight functions. Following the notations introduced above, we call a cylinder set $(a_1, a_2, \dots, a_n)_S$ *?-inclusive* if $a_i =?$ for at least one $i \in \{1, \dots, n\}$. We envisage our weight function $w(\mu)$ to be a linear combination

$$w(\mu) = \sum_{i=1}^{s} c_i \mu(\mathcal{C}_i) \tag{4.1}$$

of cylinder sets $\mathcal{C}_1, \dots, \mathcal{C}_s$, each of which is ?-inclusive, with $c_1, \dots, c_s$ being real constants (that are functions of the parameters p and q), and we want it to satisfy an inequality of the form

$$w(\widehat{\mathrm{F}}_{p,q}\, \mu) \leqslant w(\mu) - \sum_{i=1}^{s'} c'_i \mu(\mathcal{C}'_i), \tag{4.2}$$

where $c'_1, \dots, c'_{s'}$ are non-negative real constants (and are, once again, functions of p and q) and $\mathcal{C}'_1, \dots, \mathcal{C}'_{s'}$ are, once again, ?-inclusive cylinder sets. When μ is stationary for $\widehat{\mathrm{F}}_{p,q}$, we have $w(\widehat{\mathrm{F}}_{p,q}\,\mu) = w(\mu)$, which then yields $\sum_{i=1}^{s'} c'_i \mu(\mathcal{C}'_i) = 0$. The constants $c'_1, \ldots, c'_{s'}$ are such that, when $p+q>0$, we can find a *non-empty* $P \subset \{1, \ldots, s'\}$ such that $c'_i > 0$ whenever $i \in P$. This, in turn, yields $\mu(\mathcal{C}'_i) = 0$ for every $i \in P$ whenever μ is stationary and $p+q>0$. From these, we infer, due to the good choice of the constants c_i made in the weight function in (4.1), that $\mu(?) = 0$.

We outline here a summary of how we proceed in the rest of §4. Keeping in mind the goal of showing $\mu(?) = 0$ for every probability distribution μ that is stationary for the envelope PCA $\widehat{\mathrm{F}}_{p,q}$, we attempt to

1. begin with an initial, reasonable guess for the weight function,
2. parallelly, come up with a suitable inequality (or equality) that is satisfied by this initial expression for the weight function,
3. take note of the terms on the right side of this inequality (or equality) that are non-negative and therefore need to be dealt with,
4. take note of the terms on the right side of this inequality (or equality) that are non-positive and have some power in negating the above-mentioned non-negative terms,
5. accordingly, introduce an adjustment into the initial weight function,
6. write down, possibly with *several* simplifications, how the above-mentioned adjustment affects the weight function inequality (or equality), and see if the updated weight function inequality (or equality) has the desired form as stated in (4.2),
7. declare the current weight function as the final, desired weight function if the answer to the question in (6) is a yes,
8. otherwise, repeat the above procedure from (3) onward.

We conclude this introductory part of §4 by stating the final expression for the weight function that we deduce in the sequel:

$$\begin{aligned} w_3(\mu) = {} & (1-p^2-pq-q)\mu(?) + 2\mu(0?) - \mu(?0?) + 2r(1-p^2)\mu(100?) - 2pr\{\mu(1?) + \mu(10?)\} \\ & - 2p^2 r\{\mu(1??) + \mu(1?0?) + \mu(10??)\} - 4r\mu(1?01) - 2p^2 r\{\mu(1?00) + \mu(10?0)\}. \end{aligned} \tag{4.3}$$

4.1 Computation of the probabilities of various cylinder sets under the pushforward measure induced by the action of $\widehat{\mathrm{F}}_{p,q}$

Recall from §2.2 the definition of the pushforward measure $\widehat{\mathrm{F}}_{p,q}\,\mu$. Ultimately, we would need to compute $\widehat{\mathrm{F}}_{p,q}\,\mu(\mathcal{C})$ for every cylinder set $\mathcal{C}$ that shows up in the expression for the final weight function in (4.3). However, in §4.1, we compute only the first few (i.e. the pushforward measures for $(?)$, $(0?)$, $(?0?)$, $(100?)$) as these are the only cylinder sets that show up in the expression for our initial weight function (see (4.10)). The rest are computed as and when required (during the subsequent steps in which the weight function is appropriately tweaked to satisfy the desired criterion).

For *any* translation-invariant and reflection-invariant probability measure μ belonging to $\mathbb{D}$ (recall from §3), using (2.8), we have

$$\widehat{\mathrm{F}}_{p,q}\,\mu(?) = r\mu(\widehat{***}). \tag{4.4}$$

In the deduction of $\widehat{\mathrm{F}}_{p,q}\,\mu(\mathcal{C})$ where $\mathcal{C}$ is one of $0?$, $?0?$ and $100?$, we make use (2.6), (2.7) and (2.8). Since these computations involve rather similar arguments, we explain in detail only two of them.

To compute $\widehat{\mathrm{F}}_{p,q}\,\mu(0?)$, we note that for $(\widehat{\mathrm{F}}_{p,q}\,\eta(0), \widehat{\mathrm{F}}_{p,q}\,\eta(1))$ to equal $(0?)$ for some $\eta \in \hat{\mathcal{A}}^{\mathbb{Z}}$, we require $(\eta(1), \eta(2), \eta(3)) \in \widehat{***}$, and the event $\widehat{\mathrm{F}}_{p,q}\,\eta(1) = ?$ occurs with probability r. If $\eta(0) = \eta(1) = \eta(2) = 0$, forcing $\eta(3) = ?$, the event $\widehat{\mathrm{F}}_{p,q}\,\eta(0) = 0$ happens with probability p; if $\eta(0)=0$ and $(\eta(1),\eta(2),\eta(3)) \in \widehat{***} \setminus \{(00?)\}$, then again the event $\widehat{\mathrm{F}}_{p,q}\,\eta(0)=0$ happens with probability p as $(\eta(0),\eta(1),\eta(2)) \in \widehat{***}$ in this case. If $\eta(0)=?$, the event $\widehat{\mathrm{F}}_{p,q}\,\eta(0)=0$ happens with probability p, and if $\eta(0)=1$, the event $\widehat{\mathrm{F}}_{p,q}\,\eta(0)=0$ happens with probability $(1-q)$. Combining all, we have

$$\widehat{\mathrm{F}}_{p,q}\,\mu(0?) = pr\mu(\{0,?\}\widehat{***}) + (1-q)r\mu(1\widehat{***}). \tag{4.5}$$

Arguing likewise, we obtain

$$\begin{aligned}\widehat{\mathrm{F}}_{p,q}\,\mu(100?) &= (1-p)p^2r\mu(000\widehat{***}) + qp^2r[\mu(\widehat{**\widehat{*}**}) + \mu(1\{0,?\}^2\widehat{***})]\\ &\quad + q(1-q)pr\mu(1\{0,?\}\widehat{***}) + q(1-q)^2r\mu(1\widehat{***}).\end{aligned} \tag{4.6}$$

We outline the argument to deduce $\widehat{\mathrm{F}}_{p,q}\,\mu(?0?)$. For $(\widehat{\mathrm{F}}_{p,q}\,\eta(0),\widehat{\mathrm{F}}_{p,q}\,\eta(1),\widehat{\mathrm{F}}_{p,q}\,\eta(2))$ to equal $(?0?)$ for some $\eta \in \hat{\mathcal{A}}^{\mathbb{Z}}$, we require $(\eta(0),\eta(1),\eta(2)) \in \widehat{***}$ and $(\eta(2),\eta(3),\eta(4)) \in \widehat{***}$. Therefore, $(\eta(0),\eta(1)) \in \{0,?\}^2$, but we must avoid $\eta(0)=\eta(1)=\eta(2)=0$. Thus $(\eta(0),\eta(1),\eta(2),\eta(3),\eta(4))$ must belong to the set $\{0,?\}^2\widehat{***} \setminus 000\widehat{**}$. Note that this forces $(\eta(1),\eta(2),\eta(3)) \in \{0,?\}^3$, so that $\widehat{\mathrm{F}}_{p,q}\,\eta(1)=0$ happens with probability p. Each of the events $\widehat{\mathrm{F}}_{p,q}\,\eta(0)=?$ and $\widehat{\mathrm{F}}_{p,q}\,\eta(2)=?$ occurs with probability r. Combining all, we have

$$\widehat{\mathrm{F}}_{p,q}\,\mu(?0?) = pr^2[\mu(\{0,?\}^2\widehat{***}) - \mu(000\widehat{**})]. \tag{4.7}$$

4.2 Important identities used in the derivation of the weight function

We use Tables 1, 2, 3 and 4 to illustrate the derivation of a few identities. In each

-2	-1	0	1	2
	1	?	?	?
	1	?	?	0
	1	?	0	?
	1	?	0	0
	?	?	?	
	?	?	0	
0	0	?		
?	0	?		
?	?	?	1	
0	?	?	1	
1	0	?	?	
1	0	?	0	
	1	?	?	1
	1	?	0	1
	1	?	1	
1	?	?	1	
1	0	?	1	

Table 1: Decomposition of ? to establish (4.8).

-1	0	1	2
?	?	?	?
?	?	?	0
?	?	0	?
?	?	0	0
0	?	?	?
0	?	?	0
0	?	0	?
0	?	0	0
1	?	?	?
1	?	?	0
1	?	?	1
1	?	0	?
1	?	0	0
1	?	0	1
	?	1	
?	?	?	1
0	?	?	1
?	?	0	1
0	?	0	1

Table 2: Decomposition of ? to establish (4.9)

-1	0	1	2
?	0	?	?
?	0	?	0
0	0	?	?
0	0	?	0
1	0	?	?
1	0	?	0
?	0	?	1
0	0	?	1
1	0	?	1

Table 3: Decomposition of $0?$ to establish (4.9)

-1	0	1	2
?	0	0	?
0	0	0	?
1	0	0	?

Table 4: Decomposition of $00?$ to establish (4.9)

of these tables, the first row indicates the indices of the coordinates in $\mathbb{Z}$, and the rows that follow represent events involving cylinder sets. To elucidate, in Table 1, the second and third rows represent respectively the events $(\eta(-1),\eta(0),\eta(1),\eta(2)) = (1,?,?,?)$ and

$(\eta(-1), \eta(0), \eta(1), \eta(2)) = (1, ?, ?, 0)$, and it is immediate that these two events are disjoint since they disagree on the symbol that occupies the coordinate 2. In fact, for any two distinct rows that are inside the same table, it can be seen that the corresponding events are mutually exclusive. To give the reader an understanding of how we make use of these tables, note that the union of the pairwise disjoint events listed in Table 1 forms a subset of the event $\eta(0) = ?$. Thus, Table 1, along with the previously stated assumption that μ is translation-invariant and reflection-invariant (justified by Lemma 4.1), allows us to write

$$\begin{aligned}
\mu(?) &= \mu(1???) + \mu(1??0) + \mu(1?0?) + \mu(1?00) + \mu(???) + \mu(??0) + \mu(00?) + \mu(?0?) \\
&\quad + \mu(???1) + \mu(0??1) + \mu(10??) + \mu(10?0) + \mu(1??1) + \mu(1?01) + \mu(1?1) + \mu(1??1) \\
&\quad + \mu(10?1) \\
&= \underbrace{\mu(1???) + \mu(1??0) + \mu(1?0?) + \mu(1?00) + \mu(10??) + \mu(10?0)}_{(1)} + \underbrace{\mu(???1) + \mu(0??1)}_{(2)} \\
&\quad + \underbrace{\mu(???) + \mu(??0) + \mu(00?) + \mu(?0?)}_{(3)} + 2\mu(1??1) + \underbrace{\mu(1?01) + \mu(10?1)}_{(4)} + \mu(1?1) \\
&= \underbrace{\mu(1\widehat{***}) - \mu(100?)}_{\text{rewriting (1)}} + \underbrace{\mu(\widehat{***}1) - \mu(?001) - \mu(??01) - \mu(0?01) - \mu(?0?1) - \mu(00?1)}_{\text{rewriting (2)}} \\
&\quad + \underbrace{\mu(\widehat{***}) - \mu(0??) - \mu(0?0) - \mu(?00)}_{\text{rewriting (3)}} + 2\mu(1??1) + \underbrace{2\mu(1?01)}_{\text{reflection-invariance on (4)}} + \mu(1?1) \\
&= \mu(\widehat{***}) - \mu(0??) - \mu(0?0) - \mu(?00) + \underbrace{\mu(1\widehat{***}) + \mu(\widehat{***}1) - \big[\mu(100?) + \mu(?001)\big]}_{\text{from the rewritten (1) and (2) in the previous step}} \\
&\quad \underbrace{- \big[\mu(??01) + \mu(0?01)\big] - \big[\mu(?0?1) + \mu(00?1)\big]}_{\text{from the rewritten (2) in the previous step}} + 2\mu(1??1) + \mu(1?1) + 2\mu(1?01) \\
&= \mu(\widehat{***}) - \mu(0??) - \mu(0?0) - \mu(?00) + \underbrace{2\mu(1\widehat{***}) - 2\mu(100?)}_{\text{reflection-invariance}} - \big[\mu(??01) + \mu(0?01)\big] \\
&\quad - \big[\mu(?0?1) + \mu(00?1)\big] + 2\mu(1??1) + \mu(1?1) + 2\mu(1?01) \\
&= \mu(\widehat{***}) - \mu(0??) - \mu(0?0) - \mu(?00) + 2\mu(1\widehat{***}) - 2\mu(100?) - \big[\mu(?01) - \mu(1?01)\big] \\
&\quad - \big[\mu(0?1) - \mu(10?1)\big] + 2\mu(1??1) + \mu(1?1) + 2\mu(1?01) \\
&= \mu(\widehat{***}) - \big[\mu(0??) + \mu(0?0) + \mu(0?1)\big] - \big[\mu(?00) + \mu(?01)\big] + 2\mu(1\widehat{***}) - 2\mu(100?) \\
&\quad + 2\mu(1??1) + \mu(1?1) + 4\mu(1?01) \quad \text{(grouping the terms judiciously)} \\
&= \mu(\widehat{***}) - \mu(0?) - \big[\mu(?0) - \mu(?0?)\big] + 2\mu(1\widehat{***}) - 2\mu(100?) + 2\mu(1??1) + \mu(1?1) \\
&\quad + 4\mu(1?01) \\
&= \mu(\widehat{***}) - 2\mu(0?) + \mu(?0?) - 2\mu(100?) + 2\mu(1\widehat{***}) + 2\mu(1??1) + \mu(1?1) + 4\mu(1?01).
\end{aligned} \tag{4.8}$$

Likewise, Tables 2, 3 and 4 together yield

$$\begin{aligned}
&\mu(?) + \mu(0?) + \mu(00?) \\
&= \mu(????) + \mu(???0) + \mu(??0?) + \mu(??00) + \mu(0???) + \mu(0??0) + \mu(0?0?) \\
&\quad + \mu(0?00) + \mu(1???) + \mu(1??0) + \mu(1??1) + \mu(1?0?) + \mu(1?00) + \mu(1?01) + \mu(?1) \\
&\quad + \mu(???1) + \mu(0??1) + \mu(??01) + \mu(0?01) + \mu(?0??) + \mu(?0?0) + \mu(00??) + \mu(00?0) \\
&\quad + \mu(10??) + \mu(10?0) + \mu(?0?1) + \mu(00?1) + \mu(10?1) + \mu(?00?) + \mu(000?) + \mu(100?) \\
&= \big[\mu(????) + \mu(???0) + \mu(??0?) + \mu(??00) + \mu(?0??) + \mu(?0?0) + \mu(?00?)\big] + \big[\mu(0???) \\
&\quad + \mu(0??0) + \mu(0?0?) + \mu(0?00) + \mu(00??) + \mu(00?0) + \mu(000?)\big] + \big[\mu(1???) + \mu(1??0)
\end{aligned}$$

$$
\begin{aligned}
&+ \mu(1?0?) + \mu(1?00) + \mu(10??) + \mu(10?0) + \mu(100?)\big] + \big[\mu(???1) + \mu(0??1) + \mu(??01)\\
&+ \mu(0?01) + \mu(?0?1) + \mu(00?1)\big] + \mu(1??1) + \mu(1?01) + \mu(?1) + \mu(10?1)\\
&\quad \text{(grouping the terms judiciously)}\\
&= \mu(?\widehat{* * *}) + \mu(0\widehat{* * *}) + \mu(1\widehat{* * *}) + \mu(\widehat{* * * 1}) - \mu(?001) + \mu(1??1) + \mu(1?) + 2\mu(1?01)\\
&= \mu(\{0,?\}\widehat{* * *}) + 2\mu(1\widehat{* * *}) - \mu(100?) + \mu(1?) + 2\mu(1?01) + \mu(1??1). \qquad (4.9)
\end{aligned}
$$

4.3 The first step of composing the weight function

We start by defining

$$
w_0(\mu) = \mu(?) + 2\mu(0?) - \mu(?0?) + 2\mu(100?). \tag{4.10}
$$

It is not straightforward to explain our intuition behind setting w_0 in (4.10) as our 'initial' choice of weight function (after which it gets tweaked and adjusted in several steps described in the sequel to yield the final weight function). Since we are ultimately interested in $\mu(?)$ when μ is stationary, and since each $\mathcal{C}_i$ in (4.1) is ?-inclusive, it is not too far-fetched to entertain the possibility of starting with $\mathcal{C}_1 = (?)_{\{0\}}$ (i.e. the cylinder set in which ? occupies the origin). Thus, the right side of (4.2) contains $\mu(?)$, while the left contains $\widehat{\mathrm{F}}_{p,q}\,\mu(?) = r\mu(\widehat{* * *})$. When p and q are both small (intuitively, our task of showing $\mu(?) = 0$ for μ stationary ought to become harder the smaller $p + q$ gets), $r\mu(\widehat{* * *})$ is nearly equal to $\mu(\widehat{* * *})$. The appearance of $2\mu(0?)$ in the right side of (4.2) implies, from (4.5), that $2(1-q)r\mu(1\widehat{* * *})$ appears in the left side of (4.2), and when p and q are both small, this is nearly the same as $2\mu(1\widehat{* * *})$. From (4.8), we see that $\mu(?) + 2\mu(0?) - \mu(?0?) + 2\mu(100?)$ serves as an upper bound for $\mu(\widehat{* * *}) + 2\mu(1\widehat{* * *})$. All these provide ample justification as to why we start with w_0 in (4.10) as our initial weight function.

Substituting the expressions from (4.4), (4.5), (4.7) and (4.6) into (4.10), we get

$$
\begin{aligned}
w_0(\widehat{\mathrm{F}}_{p,q}\,\mu) =\ & \widehat{\mathrm{F}}_{p,q}\,\mu(?) + 2\,\widehat{\mathrm{F}}_{p,q}\,\mu(0?) - \widehat{\mathrm{F}}_{p,q}\,\mu(?0?) + 2\,\widehat{\mathrm{F}}_{p,q}\,\mu(100?)\\
=\ & r\mu(\widehat{* * *}) + 2\left\{pr\mu(\{0,?\}\widehat{* * *}) + (1-q)r\mu(1\widehat{* * *})\right\} - pr^2[\mu(\{0,?\}^2\widehat{* * *})\\
& - \mu(000\widehat{**})] + 2\big\{(1-p)p^2r\mu(000\widehat{* * *}) + qp^2r[\mu(\widehat{* * * * * *}) + \mu(1\{0,?\}^2\widehat{* * *})]\\
& + q(1-q)pr\mu(1\{0,?\}\widehat{* * *}) + q(1-q)^2r\mu(1\widehat{* * *})\big\}\\
=\ & r\mu(\widehat{* * *}) + 2\left\{pr\mu(\{0,?\}\widehat{* * *}) + (1-q)r\mu(1\widehat{* * *})\right\} - pr^2[\mu(\{0,?\}^2\widehat{* * *})\\
& - \mu(000\widehat{**})] + 2\big\{(1-p)p^2r\mu(000\widehat{* * *}) + qp^2r[\underbrace{\mu(\{0,?\}^3\widehat{* * *}) - \mu(000\widehat{* * *})}_{\text{splitting } \mu(\widehat{******})}\\
& + \mu(1\{0,?\}^2\widehat{* * *})] + q(1-q)pr\mu(1\{0,?\}\widehat{* * *}) + q(1-q)^2r\mu(1\widehat{* * *})\big\}\\
=\ & r\mu(\widehat{* * *}) + 2\left\{pr\mu(\{0,?\}\widehat{* * *}) + (1-q)r\mu(1\widehat{* * *})\right\} - pr^2[\mu(\{0,?\}^2\widehat{* * *})\\
& - \mu(000\widehat{**})] + 2\big\{(1-p)p^2r\mu(000\widehat{* * *}) + qp^2r\big[\underbrace{\mu(\{0,?\}^2\widehat{* * *})}_{\text{adding } \mu(\{0,?\}^3\widehat{***}) \text{ and } \mu(1\{0,?\}^2\widehat{***})}\\
& - \mu(000\widehat{* * *})\big] + q(1-q)pr\mu(1\{0,?\}\widehat{* * *}) + q(1-q)^2r\mu(1\widehat{* * *})\big\}\\
=\ & r\mu(\widehat{* * *}) + 2\left\{pr\mu(\{0,?\}\widehat{* * *}) + (1-q)r\mu(1\widehat{* * *})\right\} - pr^2[\mu(\{0,?\}^2\widehat{* * *})\\
& - \mu(000\widehat{**})] + 2\big\{(1-p)p^2r\mu(000\widehat{* * *}) + qp^2r\big[\underbrace{\mu(\{0,?\}\widehat{* * *}) - \mu(1\{0,?\}\widehat{* * *})}_{\text{splitting } \mu(\{0,?\}^2\widehat{***})}\\
& - \mu(000\widehat{* * *})\big] + q(1-q)pr\mu(1\{0,?\}\widehat{* * *}) + q(1-q)^2r\mu(1\widehat{* * *})\big\}\\
=\ & r\mu(\widehat{* * *}) + 2\left\{pr\mu(\{0,?\}\widehat{* * *}) + (1-q)r\mu(1\widehat{* * *})\right\} - pr^2[\mu(\{0,?\}^2\widehat{* * *})\\
& - \mu(000\widehat{**})] + 2\big\{(1-p)p^2r\mu(000\widehat{* * *}) + qp^2r\big[\underbrace{\mu(\widehat{* * *}) - \mu(1\widehat{* * *})}_{\text{splitting } \mu(\{0,?\}\widehat{***})}\\
& - \mu(1\{0,?\}\widehat{* * *}) - \mu(000\widehat{* * *})\big] + q(1-q)pr\mu(1\{0,?\}\widehat{* * *})
\end{aligned}
$$

$$
\begin{aligned}
&+ q(1-q)^2 r\mu(\widehat{1***})\big\} \\
= \quad & \underbrace{(r+2qp^2r)\mu(\widehat{***})}_{\text{combining the terms involving } \mu(\widehat{***})} \qquad + pr\mu(\{0,?\}\widehat{***}) \\
&+ \left\{ pr\mu(\{0,?\}\widehat{***}) \underbrace{+2q(1-q)pr\mu(1\{0,?\}\widehat{***}) - 2qp^2r\mu(1\{0,?\}\widehat{***})}_{\text{combining the terms involving } \mu(1\{0,?\}\widehat{***})} \right\} \\
&- pr^2[\mu(\{0,?\}^2\widehat{***}) - \mu(000\widehat{**})] \\
&+ \underbrace{\big\{2(1-p)p^2r\mu(000\widehat{***}) - 2qp^2r\mu(000\widehat{***})\big\}}_{\text{combining terms involving } \mu(000\widehat{***})} \\
&\underbrace{\big\{2q(1-q)^2r\mu(\widehat{1***}) - 2qp^2r\mu(\widehat{1***}) + 2(1-q)r\mu(\widehat{1***})\big\}}_{\text{combining the terms involving } \mu(\widehat{1***})} \\
= {} & (r+2qp^2r)\mu(\widehat{***}) + pr\mu(\{0,?\}\widehat{***}) + \big\{ \underbrace{pr\mu(\{0,?\}^2\widehat{***}) + pr\mu(1\{0,?\}\widehat{***})}_{\text{splitting } \mu(\{0,?\}\widehat{***})} \\
&+ 2qpr(1-q-p)\mu(1\{0,?\}\widehat{***})\big\} - pr^2\mu(\{0,?\}^2\widehat{***}) + pr^2\mu(000\widehat{**}) \\
&+ 2p^2r(1-p-q)\mu(000\widehat{***}) + 2r\big\{q(1-q)^2 - qp^2 + (1-q)\big\}\mu(\widehat{1***}) \\
= {} & (r+2qp^2r)\mu(\widehat{***}) + pr\mu(\{0,?\}\widehat{***}) + \ (pr+2qpr^2)\mu(1\{0,?\}\widehat{***}) \\
&+ \underbrace{pr\mu(\{0,?\}^2\widehat{***}) - pr^2\mu(\{0,?\}^2\widehat{***})}_{\text{combining terms involving } \mu(\{0,?\}^2\widehat{***})} + pr^2\mu(000\widehat{**}) + 2p^2r^2\mu(000\widehat{***}) \\
&+ 2r\left\{q + q^3 - 2q^2 - qp^2 + 1 - q\right\}\mu(\widehat{1***}) \\
= {} & (r+2qp^2r)\mu(\widehat{***}) + pr\mu(\{0,?\}\widehat{***}) + (pr+2qpr^2)\mu(1\{0,?\}\widehat{***}) \\
&+ pr(p+q)\mu(\{0,?\}^2\widehat{***}) + pr^2\mu(000\widehat{**}) + 2p^2r^2\mu(000\widehat{***}) \\
&+ 2r\left\{1 + q^3 - 2q^2 - qp^2\right\}\mu(\widehat{1***}). \qquad (4.11)
\end{aligned}
$$

Next, applying

1. the identity

$$
\mu(\widehat{***}) + 2\mu(\widehat{1***}) = \mu(?) + 2\mu(0?) - \mu(?0?) + 2\mu(100?) - 2\mu(1??1) - \mu(1?1) - 4\mu(1?01), \qquad (4.12)
$$

   obtained by rearranging the terms of (4.8),

2. and the identity

$$
\mu(\{0,?\}\widehat{***}) = \mu(?) + \mu(0?) + \mu(00?) - 2\mu(\widehat{1***}) + \mu(100?) - \mu(1?) - 2\mu(1?01) - \mu(1??1), \qquad (4.13)
$$

   obtained by rearranging the terms of (4.9), we transform (4.11) as follows:

$$
\begin{aligned}
& w_0(\widehat{\mathrm{F}}_{p,q}\,\mu) = (r+2qp^2r)\mu(\widehat{***}) + 2r\{1+q^3-2q^2-qp^2\}\mu(\widehat{1***}) + pr\mu(\{0,?\}\widehat{***}) \\
&\quad + (pr+2qpr^2)\mu(1\{0,?\}\widehat{***}) + pr(p+q)\mu(\{0,?\}^2\widehat{***}) + pr^2\mu(000\widehat{**}) \\
&\quad + 2p^2r^2\mu(000\widehat{***}) \\
= {} & \underbrace{\{r\mu(\widehat{***}) + 2r\mu(\widehat{1***})\}}_{\text{to be bound using (4.12)}} + 2qp^2r\mu(\widehat{***}) + 2r\{q^3-2q^2-qp^2\}\mu(\widehat{1***}) \underbrace{+pr\mu(\{0,?\}\widehat{***})}_{\text{to be bound using (4.13)}} \\
&\quad + (pr+2qpr^2)\mu(1\{0,?\}\widehat{***}) + pr(p+q)\mu(\{0,?\}^2\widehat{***}) + pr^2\mu(000\widehat{**}) + 2p^2r^2\mu(000\widehat{***}) \\
= {} & \underbrace{r\{\mu(?) + 2\mu(0?) - \mu(?0?) + 2\mu(100?) - 2\mu(1??1) - \mu(1?1) - 4\mu(1?01)\}}_{\text{(1) -- bound obtained from (4.12)}} + 2qp^2r\mu(\widehat{***}) \\
&\quad + 2r\{q^3-2q^2-qp^2)\}\mu(\widehat{1***}) + \underbrace{pr\{\mu(?) + \mu(0?) + \mu(00?) - 2\mu(\widehat{1***}) + \mu(100?) - \mu(1?)\}}_{\text{(2) -- bound obtained from (4.13), continued next line within underbrace}}
\end{aligned}
$$

$$
\begin{aligned}
&\underbrace{-2\mu(1?01)-\mu(1??1)\}}_{\text{(2), continued from above}}+(pr+2qpr^2)\mu(1\{0,?\}\widehat{***})+pr(p+q)\mu(\{0,?\}^2\widehat{***})\\
&+pr^2\mu(000\widehat{**})+2p^2r^2\mu(000\widehat{***})\\
=\ &\underbrace{(r+pr)\mu(?)+(2r+pr)\mu(0?)-r\mu(?0?)+pr\mu(00?)+(2r+pr)\mu(100?)-(2r+pr)\mu(1??1)}_{\text{combining terms from the bounds (1) and (2)}}\\
&\underbrace{-r\mu(1?1)-(4r+2pr)\mu(1?01)-pr\mu(1?)}_{\text{combining terms from the bounds (1) and (2)}}+2r\{q^3-2q^2-qp^2-p\}\mu(1\widehat{***})+2qp^2r\mu(\widehat{***})\\
&+(pr+2qpr^2)\mu(1\{0,?\}\widehat{***})+pr(p+q)\mu(\{0,?\}^2\widehat{***})+pr^2\mu(000\widehat{**})+2p^2r^2\mu(000\widehat{***})\\
=\ &\Big[\underbrace{\mu(?)}+(r+pr-1)\mu(?)\Big]+\Big[\underbrace{2\mu(0?)}+(2r+pr-2)\mu(0?)\Big]+\Big[\underbrace{-\mu(?0?)}-(r-1)\mu(?0?)\Big]\\
&+\Big[\underbrace{2\mu(100?)}+(2r+pr-2)\mu(100?)\Big]+pr\mu(00?)-(2r+pr)\mu(1??1)-r\mu(1?1)\\
&-(4r+2pr)\mu(1?01)-pr\mu(1?)+2r\{q^3-2q^2-qp^2-p\}\mu(1\widehat{***})+2qp^2r\mu(\widehat{***})\\
&+(pr+2qpr^2)\mu(1\{0,?\}\widehat{***})+pr(p+q)\mu(\{0,?\}^2\widehat{***})+pr^2\mu(000\widehat{**})+2p^2r^2\mu(000\widehat{***})\\
=\ &\underbrace{w_0(\mu)}_{(3)}-[p(1-r)+q]\mu(?)-[p(1-r)+q]\mu(0?)+\underbrace{(r-1)\mu(0?)-(r-1)\mu(?0?)}_{(4)}+pr\mu(00?)\\
&+(pr-2p-2q)\mu(100?)-(2r+pr)\mu(1??1)-r\mu(1?1)-(4r+2pr)\mu(1?01)-pr\mu(1?)\\
&+2r\{q^3-2q^2-qp^2-p\}\mu(1\widehat{***})+2qp^2r\mu(\widehat{***})+(pr+2qpr^2)\mu(1\{0,?\}\widehat{***})\\
&+pr(p+q)\mu(\{0,?\}^2\widehat{***})+pr^2\mu(000\widehat{**})+2p^2r^2\mu(000\widehat{***})\\
&\text{((3) obtained by combining underbraced terms from the previous step, \& using (4.10))}\\
=\ &w_0(\mu)-[p(1-r)+q]\mu(?)-[p(1-r)+q]\mu(0?)+\underbrace{(r-1)\mu(10?)+(r-1)\mu(00?)}_{\text{rewriting (4)}}+pr\mu(00?)\\
&-[p(2-r)+2q]\mu(100?)-(2r+pr)\mu(1??1)-r\mu(1?1)-(4r+2pr)\mu(1?01)-pr\mu(1?)\\
&+2r\{q^3-2q^2-qp^2-p\}\mu(1\widehat{***})+2qp^2r\mu(\widehat{***})+(pr+2qpr^2)\mu(1\{0,?\}\widehat{***})\\
&+pr(p+q)\mu(\{0,?\}^2\widehat{***})+pr^2\mu(000\widehat{**})+2p^2r^2\mu(000\widehat{***})\\
=\ &w_0(\mu)\underbrace{-[p(1-r)+q]\mu(?)-[p(1-r)+q]\mu(0?)}_{(5)}-(p+q)\mu(10?)\underbrace{-[p(1-r)+q]\mu(00?)}_{(6)}\\
&-[p(2-r)+2q]\mu(100?)-(2r+pr)\mu(1??1)-r\mu(1?1)-(4r+2pr)\mu(1?01)-pr\mu(1?)\\
&+2r\{q^3-2q^2-qp^2-p\}\mu(1\widehat{***})+2qp^2r\mu(\widehat{***})+(pr+2qpr^2)\mu(1\{0,?\}\widehat{***})\\
&+pr(p+q)\mu(\{0,?\}^2\widehat{***})+pr^2\mu(000\widehat{**})+2p^2r^2\mu(000\widehat{***})\\
=\ &w_0(\mu)\underbrace{-[p(1-r)+q][\mu(?)+\mu(0?)+\mu(00?)]}_{\text{combining (5) and (6)}}-(p+q)\mu(10?)-[p(2-r)+2q]\mu(100?)\\
&-(2r+pr)\mu(1??1)-r\mu(1?1)-(4r+2pr)\mu(1?01)-pr\mu(1?)+2qp^2r\mu(\widehat{***})\\
&-2r(p+2q^2+qp^2-q^3)\mu(1\widehat{***})+(pr+2qpr^2)\mu(1\{0,?\}\widehat{***})+pr(p+q)\mu(\{0,?\}^2\widehat{***})\\
&+pr^2\mu(000\widehat{**})+2p^2r^2\mu(000\widehat{***}).
\end{aligned}
\tag{4.14}
$$

One may note here how (4.14) lays down the first of the stepping stones which pave the way towards an inequality that resembles (4.2). We remark here that our ultimate goal would be to transform the equality in (4.14) into an inequality of the form (4.2). However, there are, at this point, several terms in the right side of (4.14) (such as $2qp^2r\mu(\widehat{***})$, $(pr+2qpr^2)\mu(1\{0,?\}\widehat{***})$ etc.), other than $w_0(\mu)$, in which the coefficients are non-negative, and this needs to be remedied. This is what we accomplish, via several adjustments and suitable algebraic manipulations in-between, in §4.4, §4.5, §4.6 and §4.7.

In each of the steps involving the above-mentioned adjustments, the equality/ inequality (beginning from (4.14)) evolves. Suppose we have performed i adjustments so far, and the weight function currently under consideration is denoted by w_i. Let the inequality that we currently have (which could actually be an equality, in which case we refer to it as the current *weight function equality*) be given by

$$w_i(\widehat{\mathrm{F}}_{p,q}\,\mu) \leqslant w_i(\mu) - \sum_{j=1}^{s_i} \alpha_{i,j}\mu(\mathcal{E}_{i,j}), \tag{4.15}$$

where $\alpha_{i,1}, \ldots, \alpha_{i,s_i}$ are real constants and $\mathcal{E}_{i,1}, \ldots, \mathcal{E}_{i,s_i}$ are cylinder sets. Suppose the $(i+1)$-st adjustment is defined via the equation $w_{i+1}(\mu) = w_i(\mu) - \sum_{j=1}^{t_i} \beta_{i,j}\mu(\mathcal{G}_{i,j})$, in which $\beta_{i,1}, \ldots, \beta_{i,t_i}$ are real constants and $\mathcal{G}_{i,1}, \ldots, \mathcal{G}_{i,t_i}$ are cylinder sets. We can now rewrite (4.15) as follows:

$$\begin{aligned} & w_{i+1}(\widehat{\mathrm{F}}_{p,q}\,\mu) + \sum_{j=1}^{t_i} \beta_{i,j}\,\widehat{\mathrm{F}}_{p,q}\,\mu(\mathcal{G}_{i,j}) \leqslant w_{i+1}(\mu) + \sum_{j=1}^{t_i} \beta_{i,j}\mu(\mathcal{G}_{i,j}) - \sum_{j=1}^{s_i} \alpha_{i,j}\mu(\mathcal{E}_{i,j}) \\ \implies & w_{i+1}(\widehat{\mathrm{F}}_{p,q}\,\mu) \leqslant w_{i+1}(\mu) + \sum_{j=1}^{t_i} \beta_{i,j}\mu(\mathcal{G}_{i,j}) - \sum_{j=1}^{t_i} \beta_{i,j}\,\widehat{\mathrm{F}}_{p,q}\,\mu(\mathcal{G}_{i,j}) - \sum_{j=1}^{s_i} \alpha_{i,j}\mu(\mathcal{E}_{i,j}). \end{aligned} \tag{4.16}$$

This is the commonality shared by all the adjustments described in the sequel, and we refer back, several times, to (4.16) and use it to see how the inequality (or equality) evolves with each adjustment.

4.4 The second step of composing the weight function

As mentioned right after (4.14), a quick comparison of (4.14) with (4.2) reveals that the terms $2qp^2r\mu(\widehat{***})$, $(pr + 2qpr^2)\mu(1\{0,?\}\widehat{***})$, $pr(p+q)\mu(\{0,?\}^2\widehat{***})$, $pr^2\mu(000\widehat{**})$ and $2p^2r^2\mu(000\widehat{***})$ in the right side of (4.14) are all non-negative. To make sure that the final inequality is of the form given in (4.2), we have to use the existing non-positive terms in the right side of (4.14) to nullify these non-negative terms.

With the aim to tackle first the terms $(pr + 2qpr^2)\mu(1\{0,?\}\widehat{***})$ and $pr(p+q)\mu(\{0,?\}^2\widehat{***})$, and then the terms $pr^2\mu(000\widehat{**})$ and $2p^2r^2\mu(000\widehat{***})$, we consider the following first adjustment to the initial weight function in (4.10):

$$w_1(\mu) = w_0(\mu) - p(p+q)\mu(?) = w_0(\mu) - p(1-r)\mu(?). \tag{4.17}$$

Rewriting (4.14) as $w_0(\widehat{\mathrm{F}}_{p,q}\,\mu) = w_0(\mu) + A_0$, and using the idea presented in (4.16), we see that the adjustment of (4.17) would transform (4.14) into

$$\begin{aligned} w_1(\widehat{\mathrm{F}}_{p,q}\,\mu) &= w_1(\mu) + p(1-r)\mu(?) - p(1-r)\,\widehat{\mathrm{F}}_{p,q}\,\mu(?) + A_0 \\ &= w_1(\mu) + p(1-r)\mu(?) - p(p+q)\,\widehat{\mathrm{F}}_{p,q}\,\mu(?) + A_0. \end{aligned} \tag{4.18}$$

We include a brief discussion attempting to explain why we choose to subtract $p(1-r)\mu(?)$ in (4.17). At the very outset, we let the reader know that we do not claim here that this is the *only* way to begin the sequence of adjustments. Some other term out of the existing non-positive terms in the right side of (4.14) might very well have worked. However, a significant amount of work went into trying out adjustments using terms that seemed plausible enough to negate the above-mentioned non-negative terms, using the idea presented in (4.16), and the choice in (4.17) has the following advantages:

1. As seen from the much neater (4.18) and by recalling the coefficient of $\mu(?)$ in the right side of (4.14), the coefficient of $\mu(?)$ becomes $-q$ once (4.17) has been

implemented, which is still non-positive and therefore creates no problems in the subsequent steps. This has been one of our main considerations in each adjustment step: that the adjustments are chosen such that they do not upset existing non-positive terms on the right side of the weight function inequality (or equality) and turn them non-negative.

2. We now come to the term $-p(p+q)\,\widehat{\mathrm{F}}_{p,q}\,\mu(?) = -p(p+q)r\mu(\widehat{***})$ in (4.18). The simple inequality $\mu(\{0,?\}^2\widehat{***}) \leqslant \mu(\widehat{***})$ tells us that $-p(p+q)r\mu(\widehat{***})$ would suffice to fully negate the non-negative term $pr(p+q)\mu(\{0,?\}^2\widehat{***})$.
3. Note that after accomplishing the task of negating $p(p+q)\mu(\{0,?\}^2\widehat{***})$, the term $-p(p+q)r\mu(\widehat{***})$ is also able to help in partially negating $(pr+2qpr^2)\mu(1\{0,?\}\widehat{***})$ (we say "partially" because the order of magnitude of the coefficient $pr+2qpr^2$ is p when p and q are small, and this is larger than the order of magnitude of $p(p+q)r$), since $\mu(1\{0,?\}\widehat{***}) + \mu(\{0,?\}^2\widehat{***}) = \mu(\{0,?\}\widehat{***}) \leqslant \mu(\widehat{***})$.

We hope that the above discussions provide some insight to the reader as to why we found $-p(1-r)\mu(?)$ a compelling choice in (4.17).

Presenting the reader with the final weight function, with all the adjustments combined together, may come across as abrupt and opaque. On the other hand, outlining the adjustments in the same chronological order in which we ourselves thought of them while developing the weight function implies that the power of each adjustment (in negating the existing non-negative terms in the right side of the weight function inequality / equality) is limited by our ability to carry out intricate computations and algebraic manipulations and to keep track of the non-negative terms that still remain. Therefore, it is sometimes the case that we carry out two different adjustments at two different stages of development of the weight function, but some of the terms used in both these adjustments are the same (up to the cylinder sets involved, but with different coefficients). For instance, in the final adjustment, i.e. (4.49), the term $-q\mu(?)$ appears, and we may very well have combined this with $-p(1-r)\mu(?)$ of (4.17), but in doing so, we would have been unable to point out effectively to the reader how the construction truly unfolds.

Applying the adjustment described in (4.17) to (4.14), as illustrated in (4.18), we obtain:

$$\begin{aligned}
&w_1(\widehat{\mathrm{F}}_{p,q}\,\mu)\\
&= w_1(\mu) + p(1-r)\mu(?) - p(1-r)\,\widehat{\mathrm{F}}_{p,q}\,\mu(?) - [p(1-r)+q][\mu(?)+\mu(0?)+\mu(00?)]\\
&\quad - (p+q)\mu(10?) - [p(2-r)+2q]\mu(100?) - (2r+pr)\mu(1??1) - r\mu(1?1)\\
&\quad - (4r+2pr)\mu(1?01) - pr\mu(1?) + 2qp^2r\mu(\widehat{***}) - 2r(p+2q^2+qp^2-q^3)\mu(1\widehat{***})\\
&\quad + (pr+2qpr^2)\mu(1\{0,?\}\widehat{***}) + pr(p+q)\mu(\{0,?\}^2\widehat{***}) + pr^2\mu(000\widehat{**}) + 2p^2r^2\mu(000\widehat{***})\\
&= w_1(\mu) - q\mu(?) - [p(1-r)+q][\mu(0?)+\mu(00?)] - (p+q)\mu(10?) - [p(2-r)+2q]\mu(100?)\\
&\quad - (2r+pr)\mu(1??1) - r\mu(1?1) - (4r+2pr)\mu(1?01) - pr\mu(1?) + 2qp^2r\mu(\widehat{***})\\
&\quad \underbrace{+pr^2\mu(000\widehat{**}) + 2p^2r^2\mu(000\widehat{***})}_{(1)} \underbrace{-p(p+q)\,\widehat{\mathrm{F}}_{p,q}\,\mu(?) - 2r(p+2q^2+qp^2-q^3)\mu(1\widehat{***})}_{(2)}\\
&\quad \underbrace{+(pr+2qpr^2)\mu(1\{0,?\}\widehat{***}) + pr(p+q)\mu(\{0,?\}^2\widehat{***})}_{(3)}.
\end{aligned} \tag{4.19}$$

Combining the terms of (4.19) that have been highlighted by the underbraces numbered (2) and (3), using (4.4), and using the identities

1. $\mu(\widehat{***}) = \mu(1\widehat{***}) + \mu(1\{0,?\}\widehat{***}) + \mu(\{0,?\}^2\widehat{***})$,

2. $\mu(1\{0,?\}\widehat{***}) = \mu(1\widehat{***}\{0,?\}) + \mu(1000?) - \mu(1?000) = \mu(1\widehat{***}) - \mu(1\widehat{***}1) + \mu(1000?) - \mu(1?000)$,

we obtain

$$\begin{aligned}
& - p(p+q)\,\widehat{\mathrm{F}}_{p,q}\,\mu(?) - 2r(p+2q^2+qp^2-q^3)\mu(1\widehat{***}) + (pr+2qpr^2)\mu(1\{0,?\}\widehat{***}) \\
& + pr(p+q)\mu(\{0,?\}^2\widehat{***}) \\
= {} & \underbrace{-p(p+q)r\mu(\widehat{***})}_{\text{use (1)}} - 2r(p+2q^2+qp^2-q^3)\mu(1\widehat{***}) + (pr+2qpr^2)\mu(1\{0,?\}\widehat{***}) \\
& + pr(p+q)\mu(\{0,?\}^2\widehat{***}) \\
= {} & \underbrace{-pr(p+q)\mu(1\widehat{***}) - pr(p+q)\mu(1\{0,?\}\widehat{***}) - pr(p+q)\mu(\{0,?\}^2\widehat{***})}_{\text{after applying (1)}} \\
& - 2r(p+2q^2+qp^2-q^3)\mu(1\widehat{***}) + (pr+2qpr^2)\mu(1\{0,?\}\widehat{***}) + pr(p+q)\mu(\{0,?\}^2\widehat{***}) \\
= {} & \underbrace{-pr(p+q)\mu(1\widehat{***}) - 2r(p+2q^2+qp^2-q^3)\mu(1\widehat{***})}_{\text{combining the terms involving } \mu(1\widehat{***})} \\
& \underbrace{-pr(p+q)\mu(1\{0,?\}\widehat{***}) + (pr+2qpr^2)\mu(1\{0,?\}\widehat{***})}_{\text{combining the terms involving } \mu(1\{0,?\}\widehat{***})} \\
& \text{(also, cancelling out the terms involving } \mu(\{0,?\}^2\widehat{***}) \text{ by each other)} \\
= {} & -r[p(p+q) + 2(p+2q^2+qp^2-q^3)]\mu(1\widehat{***}) \underbrace{+(pr^2+2qpr^2)\mu(1\{0,?\}\widehat{***})}_{\text{use (2)}} \\
= {} & -r[p(p+q) + 2(p+2q^2+qp^2-q^3)]\mu(1\widehat{***}) \\
& \underbrace{+(pr^2+2qpr^2)[\mu(1\widehat{***}) - \mu(1\widehat{***}1) + \mu(1000?) - \mu(1?000)]}_{\text{after applying (2)}} \\
= {} & -r[p(p+q) + 2(p+2q^2+qp^2-q^3)]\mu(1\widehat{***}) + (pr^2+2qpr^2)[\mu(1\widehat{***}) - \mu(1\widehat{***}1)] \\
& + (pr^2+2qpr^2)\mu(1000?) - (pr^2+2qpr^2)\mu(1?000) \\
= {} & \underbrace{-r(p+4q^2+4qp^2-2q^3+2p^2+2q^2p)\mu(1\widehat{***}) - (pr^2+2qpr^2)\mu(1\widehat{***}1)}_{(4)\text{(we combined the terms involving } \mu(1\widehat{***}))} \\
& \underbrace{+(pr^2+2qpr^2)\mu(1000?) - (pr^2+2qpr^2)\mu(1?000)}_{(5)}.
\end{aligned} \tag{4.20}$$

Next, we combine the terms of (4.20) grouped together by the underbrace numbered (5), with the terms of (4.19) highlighted by the underbrace numbered (1), and use the identities

1. $\mu(1000?) = \mu(000?) - \mu(?000?) - \mu(0000?)$,
2. $\mu(000\widehat{***}) = \mu(000?) + \mu(0000?) + \mu(00000?) - \mu(000\widehat{**}1) - \mu(000?1)$,
3. $\mu(000\widehat{**}) = \mu(000??) + \mu(000?0) + \mu(0000?)$,
4. $\mu(000??) + \mu(000?0) = \mu(000?) - \mu(000?1)$,

to obtain:

$$\begin{aligned}
& \underbrace{(pr^2+2qpr^2)\mu(1000?)}_{\text{use (1)}} - (pr^2+2qpr^2)\mu(1?000) + pr^2\mu(000\widehat{**}) \underbrace{+2p^2r^2\mu(000\widehat{***})}_{\text{use (2)}} \\
= {} & \underbrace{(pr^2+2qpr^2)[\mu(000?) - \mu(?000?) - \mu(0000?)]}_{\text{after applying (1)}} - (pr^2+2qpr^2)\mu(1?000) + pr^2\mu(000\widehat{**}) \\
& \underbrace{+2p^2r^2[\mu(000?) + \mu(0000?) + \mu(00000?) - \mu(000\widehat{**}1) - \mu(000?1)]}_{\text{after applying (2)}}
\end{aligned}$$

$$
\begin{aligned}
&= \underbrace{(pr^2+2qpr^2)\mu(000?)}_{\text{(a)}} - (pr^2+2qpr^2)\mu(0000?) \\
&\quad - (pr^2+2qpr^2)\mu(?000?) \underbrace{-(pr^2+2qpr^2)\mu(1?000)}_{\text{(c)}} \\
&\quad + \underbrace{(pr^2+2qpr^2)\mu(000\widehat{**}) - 2qpr^2\mu(000\widehat{**})}_{\text{manipulating the coefficient of } \mu(000\widehat{**})} \underbrace{+2p^2r^2\mu(000?)}_{\text{(b)}} + 2p^2r^2[\mu(0000?)+\mu(00000?)] \\
&\quad - 2p^2r^2\mu(000\widehat{**}1) \underbrace{-2p^2r^2\mu(000?1)}_{\text{(d)}} \\
&= \underbrace{(pr^2+2qpr^2+2p^2r^2)\mu(000?)}_{\text{combining (a) and (b)}} - (pr^2+2qpr^2)\mu(0000?) - \underbrace{(pr^2+2qpr^2+2p^2r^2)\mu(1?000)}_{\text{combining (c) and (d)}} \\
&\quad + \underbrace{(pr^2+2qpr^2)\mu(000\widehat{**})}_{\text{use (3)}} - (pr^2+2qpr^2)\mu(?000?) + 2p^2r^2[\mu(0000?)+\mu(00000?)] \\
&\quad - 2p^2r^2\mu(000\widehat{**}1) - 2qpr^2\mu(000\widehat{**}) \\
&= (pr^2+2qpr^2+2p^2r^2)\mu(000?) \underbrace{-(pr^2+2qpr^2)\mu(0000?)}_{\text{(e)}} - (pr^2+2qpr^2+2p^2r^2)\mu(1?000) \\
&\quad + \underbrace{(pr^2+2qpr^2)\mu(000??) + (pr^2+2qpr^2)\mu(000?0) + (pr^2+2qpr^2)\mu(0000?)}_{\text{(f): obtained by applying (3)}} \\
&\quad - (pr^2+2qpr^2)\mu(?000?) + 2p^2r^2[\mu(0000?)+\mu(00000?)] - 2p^2r^2\mu(000\widehat{**}1) \\
&\quad - 2qpr^2\mu(000\widehat{**}) \\
&= (pr^2+2qpr^2+2p^2r^2)\mu(000?) \underbrace{-(pr^2+2qpr^2)\mu(0000?) + (pr^2+2qpr^2)\mu(0000?)}_{\text{combining (e) with the last term of (f)}} \\
&\quad - (pr^2+2qpr^2+2p^2r^2)\mu(1?000) + (pr^2+2qpr^2)\Big[\underbrace{\mu(000??)+\mu(000?0)}_{\text{use (4)}}\Big] \\
&\quad - (pr^2+2qpr^2)\mu(?000?) + 2p^2r^2[\mu(0000?)+\mu(00000?)] - 2p^2r^2\mu(000\widehat{**}1) \\
&\quad - 2qpr^2\mu(000\widehat{**}) \\
&= \underbrace{(pr^2+2qpr^2+2p^2r^2)\mu(000?)}_{\text{(g)}} - \underbrace{(pr^2+2qpr^2+2p^2r^2)\mu(1?000)}_{\text{(h)}} \\
&\quad + (pr^2+2qpr^2)[\underbrace{\mu(000?)-\mu(000?1)}_{\text{(i): obtained by applying (4)}}] - (pr^2+2qpr^2)\mu(?000?) \\
&\quad + 2p^2r^2[\mu(0000?)+\mu(00000?)] - 2p^2r^2\mu(000\widehat{**}1) - 2qpr^2\mu(000\widehat{**}) \\
&= \underbrace{2pr^2(1+p+2q)\mu(000?)}_{\text{combining (g) with 1st term of (i)}} - \underbrace{2pr^2(1+2q+p)\mu(1?000)}_{\text{combining (h) with 2nd term of (i)}} + 2p^2r^2[\mu(0000?)+\mu(00000?)] \\
&\quad - 2p^2r^2\mu(000\widehat{**}1) - pr^2(1+2q)\mu(?000?) - 2qpr^2\mu(000\widehat{**}).
\end{aligned} \tag{4.21}
$$

Applying (4.20) and (4.21) to (4.19), we obtain:

$$
\begin{aligned}
&w_1(\widehat{\mathrm{F}}_{p,q}\,\mu) = w_1(\mu) - q\mu(?) - [p(1-r)+q][\mu(0?)+\mu(00?)] - (p+q)\mu(10?) \\
&- [p(2-r)+2q]\mu(100?) - (2r+pr)\mu(1??1) - r\mu(1?1) - (4r+2pr)\mu(1?01) - pr\mu(1?) \\
&+ 2qp^2r\mu(\widehat{***}) \underbrace{-r(p+4q^2+4qp^2-2q^3+2p^2+2q^2p)\mu(1\widehat{***}) - pr^2(1+2q)\mu(1\widehat{***}1)}_{\text{terms grouped by underbrace numbered (4) of (4.20)}} \\
&\underbrace{+2pr^2(1+p+2q)\mu(000?) - 2pr^2(1+2q+p)\mu(1?000) + 2p^2r^2[\mu(0000?)+\mu(00000?)]}_{\text{terms obtained from (4.21)}} \\
&\underbrace{-2p^2r^2\mu(000\widehat{**}1) - pr^2(1+2q)\mu(?000?) - 2qpr^2\mu(000\widehat{**})}_{\text{terms obtained from (4.21)}}
\end{aligned}
$$

$$
\begin{aligned}
&= w_1(\mu) - q\mu(?) - [p(1-r)+q][\mu(0?)+\mu(00?)] - (p+q)\mu(10?) - [p(2-r)+2q]\mu(100?) \\
&\quad - (2r+pr)\mu(1??1) - r\mu(1?1) - (4r+2pr)\mu(1?01) - pr\mu(1?) \underbrace{+2qp^2r\mu(\widehat{***})} \\
&\quad - r(p+4q^2+4qp^2-2q^3+2p^2+2q^2p)\mu(1\widehat{***}) - pr^2(1+2q)\mu(1\widehat{***}1) \\
&\quad - pr^2(1+2q)\mu(?000?) - 2qpr^2\mu(000\widehat{**}) \underbrace{+2pr^2(1+p+2q)\mu(000?)} \\
&\quad - 2pr^2(1+2q+p)\mu(1?000) \\
&\quad \underbrace{+2p^2r^2[\mu(0000?)+\mu(00000?)]} - 2p^2r^2\mu(000\widehat{**}1) \\
&\quad \text{(only rearrangements done in this step).} \qquad (4.22)
\end{aligned}
$$

Again, (4.22), much like (4.14), forms part of the foundation upon which the derivation of an inequality of the form (4.2) is built. Still, there are terms in the right side of (4.22), other than $w_1(\mu)$, indicated by underbraces in the final step of the derivation of (4.22), in which the coefficients are non-negative. Therefore, further adjustments are necessary.

4.5 The third step of composing the weight function

In order to make sure that we ultimately end up with an inequality of the form given in (4.2), we have to ensure that the non-negative terms highlighted by underbraces in the final expression for (4.22) are negated using some of the existing non-positive terms in the final expression for (4.22).

The second adjustment is carried out as follows:

$$
\begin{aligned}
w_2(\mu) = w_1(\mu) - \big[&2pr\{\mu(1?)+\mu(10?)\} + 2p^2r\{\mu(1??)+\mu(1?0?)+\mu(10??)\} + 4r\mu(1?01) \\
&+ 2p\mu(100?)\big]. \qquad (4.23)
\end{aligned}
$$

Again, as stated at the beginning of §4.4, a different choice of terms may very well have worked instead of those considered in (4.23). Writing (4.22) as $w_1(\widehat{\mathrm{F}}_{p,q}\,\mu) = w_1(\mu) + A_1$, we see, using the idea presented in (4.16), that applying (4.23) will transform (4.22) into

$$
\begin{aligned}
w_2(\widehat{\mathrm{F}}_{p,q}\,\mu) = w_2(\mu) &+ \{2pr\{\mu(1?)+\mu(10?)\} + 2p^2r\{\mu(1??)+\mu(1?0?)+\mu(10??)\} \\
&+ 4r\mu(1?01) + 2p\mu(100?)\} - \{2pr\{\widehat{\mathrm{F}}_{p,q}\,\mu(1?) + \widehat{\mathrm{F}}_{p,q}\,\mu(10?)\} \\
&+ 2p^2r\{\widehat{\mathrm{F}}_{p,q}\,\mu(1??) + \widehat{\mathrm{F}}_{p,q}\,\mu(1?0?) + \widehat{\mathrm{F}}_{p,q}\,\mu(10??)\} + 4r\,\widehat{\mathrm{F}}_{p,q}\,\mu(1?01) \\
&+ 2p\,\widehat{\mathrm{F}}_{p,q}\,\mu(100?)\} + A_1. \qquad (4.24)
\end{aligned}
$$

As in §4.4, we attempt to explain to the reader some of the motivations for choosing the above adjustment. Each of the terms used in (4.23) has its own significance, and these are, of course, detailed in the computations that follow for the reader to verify, so we focus on just the reason for our choice of the term $-2pr\mu(1?)$.

The terms in the right side of (4.22) that are non-negative are $2qp^2r\mu(\widehat{***})$, $2pr^2(1+p+2q)\mu(000?)$ and $2p^2r^2[\mu(0000?)+\mu(00000?)]$. When p and q are small, the order of magnitude of the coefficient of the first of these terms is $2qp^2$, that of the second term is $2p$, and that of the third term is $2p^2$. We emphasize here, to the reader, that in our approach for constructing the weight function, we have tried to first annihilate the non-negative terms whose coefficients have higher orders of magnitude, and then addressed the non-negative terms whose coefficients have lower orders of magnitude. So, here, we first focus on $2pr^2(1+p+2q)\mu(000?)$, which can be thought of as being approximately equal to $2p\mu(000?)$ when p and q are small.

From (4.28), we know that one of the terms in the expansion of $\widehat{\mathrm{F}}_{p,q}\,\mu(1?)$ is $r^2\mu(000?)$. Therefore, $-2pr\,\widehat{\mathrm{F}}_{p,q}\,\mu(1?)$ in (4.24) will contribute $-2pr^3\mu(000?)$, which is approximately equal to $-2p\mu(000?)$ when p and q are small. This, then, aids in negating much of the non-negative term $2pr^2(1+p+2q)\mu(000?)$ – in fact, what is left of $2pr^2(1+p+2q)\mu(000?)$ after implementing this adjustment is a non-negative term whose coefficient is of a smaller order of magnitude.

At the same time, we note that on the right side of (4.24), the term $2pr\mu(1?)$ is being added, and we need to make sure that this does not upset any of the existing non-positive terms to the extent that they become non-negative. Note that there already exists a $-pr\mu(1?)$ on the right side of (4.22). Next, we note that

1. $-r(p+4q^2+4qp^2-2q^3+2p^2+2q^2p)\mu(1\widehat{***})$ on the right side of (4.22) supplies us with the terms $-pr\mu(1???)$, $-pr\mu(1??0)$, $-pr\mu(1?0?)$ and $-pr\mu(1?00)$,
2. the term $-(2r+pr)\mu(1??1)$ provides us with $-pr\mu(1??1)$,
3. the term $-(4r+2pr)\mu(1?01)$ provides us with $-pr\mu(1?01)$,
4. and the term $-r\mu(1?1)$ supplies us with $-pr\mu(1?1)$.

We see that $-pr\mu(1???) - pr\mu(1??0) - pr\mu(1?0?) - pr\mu(1?00) - pr\mu(1??1) - pr\mu(1?01) - pr\mu(1?1) = -pr\mu(1?)$. Thus, we have accounted for non-positive terms existing on the right side of (4.22) that are capable of negating, together, the $2pr\mu(1?)$ that gets introduced into the right side of (4.24) because of (4.23).

Once again, we hope that the above discussion is able to shed some light on why $-2pr\mu(1?)$ is a part of (4.23). Likewise, the presence of the other terms in (4.23) can be justified / motivated.

Applying the adjustment described in (4.23) to (4.22), we obtain (as shown in (4.24)):

$$
\begin{aligned}
& w_2(\widehat{\mathrm{F}}_{p,q}\,\mu) = w_2(\mu) + [2pr\{\mu(1?)+\mu(10?)\} + 2p^2r\{\mu(1??)+\mu(1?0?)+\mu(10??)\} \\
& + 4r\mu(1?01) + 2p\mu(100?)] - [2pr\{\widehat{\mathrm{F}}_{p,q}\,\mu(1?) + \widehat{\mathrm{F}}_{p,q}\,\mu(10?)\} + 2p^2r\{\widehat{\mathrm{F}}_{p,q}\,\mu(1??) \\
& + \widehat{\mathrm{F}}_{p,q}\,\mu(1?0?) + \widehat{\mathrm{F}}_{p,q}\,\mu(10??)\} + 4r\,\widehat{\mathrm{F}}_{p,q}\,\mu(1?01) + 2p\,\widehat{\mathrm{F}}_{p,q}\,\mu(100?)] - q\mu(?) \\
& - [p(1-r)+q][\mu(0?)+\mu(00?)] - (p+q)\mu(10?) \\
& - [p(2-r)+2q]\mu(100?) - (2r+pr)\mu(1??1) - r\mu(1?1) - (4r+2pr)\mu(1?01) - pr\mu(1?) \\
& + 2qp^2r\mu(\widehat{***}) - r(p+4q^2+4qp^2-2q^3+2p^2+2q^2p)\mu(1\widehat{***}) - pr^2(1+2q)\mu(1\widehat{***}1) \\
& - pr^2(1+2q)\mu(?000?) - 2qpr^2\mu(000\widehat{**}) + 2pr^2(1+p+2q)\mu(000?) \\
& - 2pr^2(1+2q+p)\mu(1?000) + 2p^2r^2[\mu(0000?)+\mu(00000?)] - 2p^2r^2\mu(000\widehat{**}1) \\
= {} & w_2(\mu) + \underbrace{2pr\{\mu(1?)+\mu(10?)\} + 2p^2r\{\mu(1??)+\mu(1?0?)+\mu(10??)\} + 4r\mu(1?01) + 2p\mu(100?)} \\
& \underbrace{-(p+q)\mu(10?) - [p(2-r)+2q]\mu(100?) - (2r+pr)\mu(1??1) - r\mu(1?1) - (4r+2pr)\mu(1?01) - pr\mu(1?)} \\
& \underbrace{-r(p+4q^2+4qp^2-2q^3+2p^2+2q^2p)\mu(1\widehat{***})} - q\mu(?) - [p(1-r)+q][\mu(0?)+\mu(00?)] \\
& + 2qp^2r\mu(\widehat{***}) - pr^2(1+2q)\mu(1\widehat{***}1) - pr^2(1+2q)\mu(?000?) - 2qpr^2\mu(000\widehat{**}) \\
& + 2pr^2(1+p+2q)\mu(000?) - 2pr^2(1+2q+p)\mu(1?000) + 2p^2r^2[\mu(0000?)+\mu(00000?)] \\
& - 2p^2r^2\mu(000\widehat{**}1) - 2pr\{\widehat{\mathrm{F}}_{p,q}\,\mu(1?) + \widehat{\mathrm{F}}_{p,q}\,\mu(10?)\} - 2p^2r\{\widehat{\mathrm{F}}_{p,q}\,\mu(1??) + \widehat{\mathrm{F}}_{p,q}\,\mu(1?0?) \\
& + \widehat{\mathrm{F}}_{p,q}\,\mu(10??)\} - 4r\,\widehat{\mathrm{F}}_{p,q}\,\mu(1?01) - 2p\,\widehat{\mathrm{F}}_{p,q}\,\mu(100?).
\end{aligned}
\tag{4.25}
$$

4.5.1 Step 1 of analyzing (4.25)

First, we combine the terms of (4.25) that have been highlighted using underbraces. To this end, we use the identities:

1. $\mu(1\widehat{***}) = \mu(1?) + \mu(10?) + \mu(100?) - \mu(1?1) - \mu(1??1) - 2\mu(1?01)$,
2. $\mu(1???) + \mu(1??0) = \mu(1??) - \mu(1??1)$,
3. $\mu(1\widehat{***}) = \mu(1???) + \mu(1??0) + \mu(1?0?) + \mu(10??) + \mu(1?00) + \mu(10?0) + \mu(100?)$,

and we obtain:

$$
\begin{aligned}
&\underbrace{-(p+q)\mu(10?)}_{(1)}\underbrace{-\{p(2-r)+2q\}\mu(100?)}_{(3)}-(2r+pr)\mu(1??1)-r\mu(1?1)\underbrace{-(4r+2pr)\mu(1?01)}_{(5)}\\
&\underbrace{-pr\mu(1?)}_{(7)}-r(p+4q^2+4qp^2-2q^3+2p^2+2q^2p)\mu(\widehat{1***})\underbrace{+2pr\mu(1?)}_{(8)}\underbrace{+2pr\mu(10?)}_{(2)}\\
&+2p^2r\{\mu(1??)+\mu(1?0?)+\mu(10??)\}\underbrace{+4r\mu(1?01)}_{(6)}\underbrace{+2p\mu(100?)}_{(4)}\\
=\;&\underbrace{-(p+q)\mu(10?)+2pr\mu(10?)}_{\text{combining (1) and (2)}}\underbrace{-\{p(2-r)+2q\}\mu(100?)+2p\mu(100?)}_{\text{combining (3) and (4)}}-(2r+pr)\mu(1??1)\\
&-r\mu(1?1)\underbrace{-(4r+2pr)\mu(1?01)+4r\mu(1?01)}_{\text{combining (5) and (6)}}\underbrace{-pr\mu(1?)+2pr\{\mu(1?)}_{\text{combining (7) and (8)}}\underbrace{-pr\mu(\widehat{1***})}_{\text{apply (1)}}\\
&-2p^2r\underbrace{\mu(\widehat{1***})}_{\text{split using (3)}}-r(4q^2+4qp^2-2q^3+2q^2p)\mu(\widehat{1***})\\
&+2p^2r\{\mu(1??)+\mu(1?0?)+\mu(10??)\}\\
=\;&[-(p+q)+2pr]\mu(10?)-(2q-pr)\mu(100?)-(2r+pr)\mu(1??1)-r\mu(1?1)\\
&-2pr\mu(1?01)+pr\mu(1?)\\
&\underbrace{-pr[\mu(1?)+\mu(10?)+\mu(100?)-\mu(1?1)-\mu(1??1)-2\mu(1?01)]}_{\text{after applying (1)}}-2p^2r\big[\mu(1???)+\mu(1??0)\\
&+\mu(1?0?)+\mu(10??)+\mu(1?00)+\mu(10?0)+\mu(100?)\big]\\
&-r(4q^2+4qp^2-2q^3+2q^2p)\mu(\widehat{1***})+2p^2r\{\mu(1??)+\mu(1?0?)+\mu(10??)\}\\
=\;&\underbrace{[-(p+q)+2pr]\mu(10?)}_{(9)}\underbrace{-(2q-pr)\mu(100?)}_{(11)}\underbrace{-(2r+pr)\mu(1??1)}_{(13)}\underbrace{-r\mu(1?1)}_{(15)}\underbrace{-2pr\mu(1?01)}_{(17)}\\
&\underbrace{+pr\mu(1?)}_{(19)}\underbrace{-pr\mu(1?)}_{(20)}\underbrace{-pr\mu(10?)}_{(10)}\underbrace{-pr\mu(100?)}_{(12)}\underbrace{+pr\mu(1?1)}_{(16)}\underbrace{+pr\mu(1??1)}_{(14)}\underbrace{+2pr\mu(1?01)}_{(18)}\\
&-2p^2r\big[\underbrace{\mu(1???)+\mu(1??0)}_{\text{apply (2)}}+\mu(1?0?)+\mu(10??)+\mu(1?00)+\mu(10?0)+\mu(100?)\big]\\
&-r(4q^2+4qp^2-2q^3+2q^2p)\underbrace{\mu(\widehat{1***})}_{\text{split using (3)}}+2p^2r\{\mu(1??)+\mu(1?0?)+\mu(10??)\}\\
=\;&\underbrace{[-(p+q)+pr]\mu(10?)}_{\text{adding (9) \& (10)}}\underbrace{-2q\mu(100?)}_{\text{adding (11) \& (12)}}\underbrace{-2r\mu(1??1)}_{\text{adding (13) \& (14)}}\underbrace{-r(1-p)\mu(1?1)}_{\text{adding (15) \& (16)}}\\
&-2p^2r[\underbrace{\mu(1??)-\mu(1??1)}_{\text{after applying (2)}}+\mu(1?0?)+\mu(10??)]-2p^2r[\mu(1?00)+\mu(10?0)+\mu(100?)]\\
&-r(4q^2+4qp^2-2q^3+2q^2p)[\mu(1???)+\mu(1??0)+\mu(1?0?)+\mu(10??)+\mu(1?00)\\
&+\mu(10?0)+\mu(100?)]+2p^2r\{\mu(1??)+\mu(1?0?)+\mu(10??)\}
\end{aligned}
$$

(also adding the terms indicated by underbraces numbered (17), (18), (19) & (20))

$$
\begin{aligned}
=\;&-[p(1-r)+q]\mu(10?)\underbrace{-2q\mu(100?)}_{(21)}\underbrace{-2r\mu(1??1)}_{(24)}-r(1-p)\mu(1?1)\underbrace{+2p^2r\mu(1??1)}_{(25)}\\
&\underbrace{-2p^2r[\mu(1??)+\mu(1?0?)+\mu(10??)]}_{(26)}\underbrace{-2p^2r[\mu(1?00)+\mu(10?0)]}_{(28)}\underbrace{-2p^2r\mu(100?)}_{(22)}\\
&-r(4q^2+4qp^2-2q^3+2q^2p)[\mu(1???)+\mu(1??0)+\mu(1?0?)+\mu(10??)\underbrace{+\mu(1?00)+\mu(10?0)}_{(29)}]
\end{aligned}
$$

$$
\begin{aligned}
&\underbrace{-r(4q^2+4qp^2-2q^3+2q^2p)\mu(100?)}_{(23)}\underbrace{+2p^2r\{\mu(1??)+\mu(1?0?)+\mu(10??)\}}_{(27)}\\
=&-[p(1-r)+q]\mu(10?)\underbrace{-[2q+r(4q^2+4qp^2-2q^3+2p^2+2q^2p)]\mu(100?)}_{\text{adding (21), (22), (23)}}\\
&\underbrace{-2r(1-p^2)\mu(1??1)}_{\text{adding (24) \& (25)}}\\
&-r(1-p)\mu(1?1)-r(4q^2+4qp^2-2q^3+2q^2p)[\mu(1???)+\mu(1??0)+\mu(1?0?)+\mu(10??)]\\
&\underbrace{-r(4q^2+4qp^2-2q^3+2p^2+2q^2p)[\mu(1?00)+\mu(10?0)]}_{\text{adding (28) \& (29)}}\quad\text{(also cancelling (26) by (27)).}
\end{aligned}
\tag{4.26}
$$

Incorporating the expression obtained in (4.26) (as indicated by the large square brackets in the expression below), we can now rewrite (4.25) as follows:

$$
\begin{aligned}
&w_2(\widehat{\mathrm{F}}_{p,q}\,\mu)\\
&=w_2(\mu)+\Bigg[-[p(1-r)+q]\mu(10?)-[2q+r(4q^2+4qp^2-2q^3+2p^2+2q^2p)]\mu(100?)\\
&-2r(1-p^2)\mu(1??1)-r(1-p)\mu(1?1)-r(4q^2+4qp^2-2q^3+2q^2p)[\mu(1???)+\mu(1??0)\\
&+\mu(1?0?)+\mu(10??)]-r(4q^2+4qp^2-2q^3+2p^2+2q^2p)[\mu(1?00)+\mu(10?0)]\Bigg]-q\mu(?)\\
&-[p(1-r)+q][\mu(0?)+\mu(00?)]+2qp^2r\mu(\widehat{***})-pr^2(1+2q)\mu(1\widehat{***}1)\\
&-pr^2(1+2q)\mu(?000?)-2qpr^2\mu(000\widehat{**})+2pr^2(1+p+2q)\mu(000?)\\
&-2pr^2(1+2q+p)\mu(1?000)+2p^2r^2[\mu(0000?)+\mu(00000?)]\\
&-2p^2r^2\mu(000\widehat{**}1)\underbrace{-2pr\{\widehat{\mathrm{F}}_{p,q}\,\mu(1?)+\widehat{\mathrm{F}}_{p,q}\,\mu(10?)\}-2p^2r\{\widehat{\mathrm{F}}_{p,q}\,\mu(1??)}\\
&\underbrace{+\widehat{\mathrm{F}}_{p,q}\,\mu(1?0?)+\widehat{\mathrm{F}}_{p,q}\,\mu(10??)\}-4r\,\widehat{\mathrm{F}}_{p,q}\,\mu(1?01)-2p\,\widehat{\mathrm{F}}_{p,q}\,\mu(100?)}.
\end{aligned}
\tag{4.27}
$$

4.5.2 Step 2 of analyzing (4.25)

In §4.5.3, we deal with the terms of (4.27) that have been grouped using underbraces. To this end, we need to compute $\widehat{\mathrm{F}}_{p,q}\,\mu(1?)$, $\widehat{\mathrm{F}}_{p,q}\,\mu(10?)$, $\widehat{\mathrm{F}}_{p,q}\,\mu(1??)$, $\widehat{\mathrm{F}}_{p,q}\,\mu(1?0?)$, $\widehat{\mathrm{F}}_{p,q}\,\mu(10??)$ and $\widehat{\mathrm{F}}_{p,q}\,\mu(1?01)$ (note that we have already computed $\widehat{\mathrm{F}}_{p,q}\,\mu(100?)$ in (4.6)). An argument similar to that adopted in (4.5) of §4.1 for computing $\widehat{\mathrm{F}}_{p,q}\,\mu(0?)$ can be used to deduce that

$$\widehat{\mathrm{F}}_{p,q}\,\mu(1?)=r^2\mu(000?)+qr\mu(\widehat{***}). \tag{4.28}$$

However, parts of the expression for each of the probabilities $\widehat{\mathrm{F}}_{p,q}\,\mu(10?)$, $\widehat{\mathrm{F}}_{p,q}\,\mu(1??)$, $\widehat{\mathrm{F}}_{p,q}\,\mu(1?0?)$, $\widehat{\mathrm{F}}_{p,q}\,\mu(10??)$ and $\widehat{\mathrm{F}}_{p,q}\,\mu(1?01)$ will deliberately *not* be made explicit, in order to keep the subsequent mathematical expressions as concise as possible.

While computing $\widehat{\mathrm{F}}_{p,q}\,\mu(10?)$, we first consider $\eta(0)=\eta(1)=\eta(2)=0$, so that $(\eta(3),\eta(4))\in\widehat{**}$, the event $\widehat{\mathrm{F}}_{p,q}\,\eta(0)=1$ happens with probability $1-p$, and the event $\widehat{\mathrm{F}}_{p,q}\,\eta(1)=0$ happens with probability p. Next, we consider $\eta(0)=1$, $\eta(1)\in\{0,?\}$ and $(\eta(2),\eta(3),\eta(4))\in\widehat{***}$, so that $\widehat{\mathrm{F}}_{p,q}\,\eta(0)=1$ happens with probability q and $\widehat{\mathrm{F}}_{p,q}\,\eta(1)=0$ happens with probability p, and finally, we consider $\eta(1)=1$ and $(\eta(2),\eta(3),\eta(4))\in\widehat{***}$, so that $\widehat{\mathrm{F}}_{p,q}\,\eta(0)=1$ happens with probability q and $\widehat{\mathrm{F}}_{p,q}\,\eta(1)=0$ happens with probability $1-q$. In each of these cases, $\widehat{\mathrm{F}}_{p,q}\,\eta(2)=?$ happens with probability r. Therefore

$$\widehat{\mathrm{F}}_{p,q}\,\mu(10?)=(1-p)pr\mu(000\widehat{**})+C_{10?}+D_{10?}\geqslant(1-p)pr\mu(000\widehat{**})+C_{10?}, \tag{4.29}$$

where

$$C_{10?} = qpr\mu(1\{0,?\}\widehat{***}) + q(1-q)r\mu(1\widehat{***}), \tag{4.30}$$

and $D_{10?}$ is the component arising from the case where $(\eta(0), \eta(1), \eta(2)) \in \widehat{***}$. Similar arguments lead to

- $\widehat{\mathrm{F}}_{p,q}\,\mu(1??) = (1-p)r^2\mu(000?\{0,?\}) + C_{1??} \geqslant (1-p)r^2\mu(000?\{0,?\});$ (4.31)
- $\widehat{\mathrm{F}}_{p,q}\,\mu(1?0?) = (1-p)r^2p\mu(000?\{0,?\}^2) + C_{1?0?} \geqslant (1-p)r^2p\mu(000?\{0,?\}^2);$ (4.32)
- $\widehat{\mathrm{F}}_{p,q}\,\mu(10??) = (1-p)pr^2\mu(000\widehat{**}\{0,?\}) + C_{10??} \geqslant (1-p)pr^2\mu(000\widehat{**}\{0,?\});$ (4.33)

in which $C_{\mathcal{C}}$, where $\mathcal{C}$ is any of the cylinder sets $(1??)$, $(1?0?)$, $(10??)$, accounts for the contribution from the cases in which $(\eta(0), \eta(1), \eta(2)) \in \hat{\mathcal{A}}^3 \setminus \{(0,0,0)\}$.

While computing $\widehat{\mathrm{F}}_{p,q}\,\mu(1?01)$, we first consider the case where $\eta(3) = \eta(4) = \eta(5) = 0$, so that $(\eta(1), \eta(2)) \in \widehat{**}$, and the event $\widehat{\mathrm{F}}_{p,q}\,\eta(2) = 0$ happens with probability p while the event $\widehat{\mathrm{F}}_{p,q}\,\eta(3) = 1$ happens with probability $1-p$. Note that in this situation, no matter what the value of $\eta(0)$ is, the event $\widehat{\mathrm{F}}_{p,q}\,\eta(0) = 1$ happens with probability q. The second possibility we take into account is where $\eta(0) = \eta(1) = \eta(2) = 0$, which then forces $\eta(3) = ?$, and $\widehat{\mathrm{F}}_{p,q}\,\eta(0) = 1$ happens with probability $1-p$. If $\eta(4) \in \{0,?\}$, the events $\widehat{\mathrm{F}}_{p,q}\,\eta(2) = 0$ and $\widehat{\mathrm{F}}_{p,q}\,\eta(3) = 1$ happen with probabilities p and q respectively, whereas if $\eta(4) = 1$, they happen with probabilities $1-q$ and q respectively. Combining all, we have

$$\begin{aligned}
&\widehat{\mathrm{F}}_{p,q}\,\mu(1?01) \\
&= qrp(1-p)\mu(\widehat{**}000) + (1-p)rpq\mu(000?\{0,?\}) + (1-p)r(1-q)q\mu(000?1) + C_{1?01} \\
&= qrp(1-p)\left\{\mu(?000) - \mu(1?000) + \mu(?0000)\right\} + (1-p)rpq\left\{\mu(000?) - \mu(000?1)\right\} \\
&\quad + (1-p)r(1-q)q\mu(000?1) + C_{1?01} \\
&= \underbrace{qrp(1-p)\mu(?000) + (1-p)rpq\mu(000?)} + qrp(1-p)\mu(?0000) \\
&\quad \underbrace{-qrp(1-p)\mu(1?000) - (1-p)rpq\mu(000?1) + (1-p)r(1-q)q\mu(000?1)} + C_{1?01} \\
&\geqslant 2(1-p)rpq\mu(000?) + (1-p)prq\mu(0000?) + (1-p)rq(r-p)\mu(000?1),
\end{aligned} \tag{4.34}$$

where we make use of the reflection-invariance of μ, and $C_{1?01}$ takes into account the contributions from the situations not considered above.

Finally, instead of writing the entire expression of (4.6) in place of $\widehat{\mathrm{F}}_{p,q}\,\mu(100?)$, we write

$$\widehat{\mathrm{F}}_{p,q}\,\mu(100?) = p^2r^2\mu(000\widehat{***}) + C_{100?}, \tag{4.35}$$

where

$$C_{100?} = qp^2r\mu(\widehat{***}) + qpr^2\mu(1\{0,?\}\widehat{***}) + qr^2(1-q+p)\mu(1\widehat{***}). \tag{4.36}$$

4.5.3 Step 3 of analyzing (4.25)

As mentioned at the start of §4.5.2, we deal with the terms grouped using underbraces in (4.27), using (4.28), (4.29), (4.31), (4.32), (4.33), (4.34), (4.35), and applying the identities

1. $\mu(000\widehat{***}) = \mu(000?) + \mu(0000?) + \mu(00000?) - \mu(000\widehat{**}1) - \mu(000?1),$
2. $\mu(000\widehat{**}) = \mu(000?) - \mu(000?1) + \mu(0000?),$
3. $\mu(000?\{0,?\}) = \mu(000?) - \mu(000?1),$
4. $\mu(000?\{0,?\}^2) = \mu(000?) - \mu(000?1) - \mu(000?\{0,?\}1),$

5. $\mu(000\widehat{**}\{0,?\}) = \mu(000?) + \mu(0000?) - \mu(000?1) - \mu(000?\{0,?\}1) - \mu(0000?1)$,

Using (4.28), (4.29), (4.31), (4.32), (4.33), (4.34) and (4.35) in the first step, and subsequently applying the identities mentioned above, we find that the sum of the terms grouped using underbraces in (4.27) can be bounded above as follows:

$$\begin{aligned}
& -2pr\{\widehat{\mathrm{F}}_{p,q}\,\mu(1?) + \widehat{\mathrm{F}}_{p,q}\,\mu(10?)\} - 2p^2r\{\widehat{\mathrm{F}}_{p,q}\,\mu(1??) + \widehat{\mathrm{F}}_{p,q}\,\mu(1?0?) + \widehat{\mathrm{F}}_{p,q}\,\mu(10??)\} \\
& -4r\,\widehat{\mathrm{F}}_{p,q}\,\mu(1?01) - 2p\,\widehat{\mathrm{F}}_{p,q}\,\mu(100?) \\
\leqslant\ & -2pr^3\mu(000?) - 2pqr^2\mu(\widehat{***}) - 2p^2r^2(1-p)\underbrace{\mu(000\widehat{**})}_{\text{use (2)}} - 2prC_{10?} \\
& -2p^2r^3(1-p)\underbrace{\mu(000?\{0,?\})}_{\text{use (3)}} \\
& -2p^3r^3(1-p)\underbrace{\mu(000?\{0,?\}^2)}_{\text{use (4)}} - 2p^3r^3(1-p)\underbrace{\mu(000\widehat{**}\{0,?\})}_{\text{use (5)}} - 8(1-p)r^2pq\mu(000?) \\
& -4(1-p)pr^2q\mu(0000?) - 4(1-p)r^2q(r-p)\mu(000?1) - 2p^3r^2\underbrace{\mu(000\widehat{***})}_{\text{use (1)}} - 2pC_{100?} \\
=\ & -2pr^3\mu(000?) - 2pqr^2\mu(\widehat{***}) - 2p^2r^2(1-p)\underbrace{[\mu(000?) - \mu(000?1) + \mu(0000?)]}_{\text{after applying (2)}} \\
& -2p^2r^3(1-p)\underbrace{[\mu(000?) - \mu(000?1)]}_{\text{after applying (3)}} - 2p^3r^3(1-p)\underbrace{[\mu(000?) - \mu(000?1) - \mu(000?\{0,?\}1)]}_{\text{after applying (4)}} \\
& -2p^3r^3(1-p)\underbrace{[\mu(000?) - \mu(000?1) - \mu(000?\{0,?\}1) + \mu(0000?) - \mu(0000?1)]}_{\text{after applying (5)}} \\
& -8(1-p)r^2pq\mu(000?) - 4(1-p)pr^2q\mu(0000?) - 4(1-p)r^2q(r-p)\mu(000?1) \\
& -2p^3r^2\underbrace{[\mu(000?) + \mu(0000?) + \mu(00000?) - \mu(000?1) - \mu(000?\{0,?\}1) - \mu(0000?1)]}_{\text{after applying (1)}} \\
& -2prC_{10?} - 2pC_{100?} \\
=\ & -[2pr^3 + 2p^2r^2(1-p) + 2p^2r^3(1-p) + 4p^3r^3(1-p) + 8(1-p)r^2pq + 2p^3r^2]\mu(000?) \\
& -[2p^2r^2(1-p) + 2p^3r^3(1-p) + 4(1-p)pr^2q + 2p^3r^2]\mu(0000?) - 2p^3r^2\mu(00000?) \\
& +[2p^2r^2(1-p) + 2p^2r^3(1-p) + 4p^3r^3(1-p) - 4(1-p)r^2q(r-p) + 2p^3r^2]\mu(000?1) \\
& +[4p^3r^3(1-p) + 2p^3r^2]\mu(000?\{0,?\}1) + [2p^3r^3(1-p) + 2p^3r^2]\mu(0000?1) \\
& -2pqr^2\mu(\widehat{***}) - 2prC_{10?} - 2pC_{100?},
\end{aligned} \tag{4.37}$$

where, in the last step, we simply add all those terms that involve $\mu(\mathcal{C})$, for $\mathcal{C}$ being any of the cylinder sets $(000?)$, $(0000?)$, $(00000?)$, $(000?1)$, $(000?\{0,?\}1)$ and $(0000?1)$. Incorporating the inequality (4.37), we now see that the equality in (4.27) is transformed into an inequality as follows (it is worthwhile to note here that we encountered only weight function *equalities* in the various steps of adjustments carried out earlier, and this is the first time we encounter a weight function *inequality*):

$$\begin{aligned}
& w_2(\widehat{\mathrm{F}}_{p,q}\,\mu) \leqslant w_2(\mu) - [p(1-r) + q]\mu(10?) - [2q + r(4q^2 + 4qp^2 - 2q^3 + 2p^2 + 2q^2p)]\mu(100?) \\
& -2r(1-p^2)\mu(1??1) - r(1-p)\mu(1?1) - r(4q^2 + 4qp^2 - 2q^3 + 2q^2p)[\mu(1???) + \mu(1??0) \\
& +\mu(1?0?) + \mu(10??)] - r(4q^2 + 4qp^2 - 2q^3 + 2p^2 + 2q^2p)[\mu(1?00) + \mu(10?0)] - q\mu(?) \\
& -[p(1-r) + q][\mu(0?) + \mu(00?)] - pr^2(1+2q)\mu(1\widehat{***}1) \\
& \underbrace{+2qp^2r\mu(\widehat{***}) - pr^2(1+2q)\mu(?000?) - 2qpr^2\mu(000\widehat{**}) + 2pr^2(1+p+2q)\mu(000?)} \\
& \underbrace{-2pr^2(1+2q+p)\mu(1?000) + 2p^2r^2[\mu(0000?) + \mu(00000?)] - 2p^2r^2\mu(000\widehat{**}1)}
\end{aligned}$$

$$\begin{aligned}
&\underbrace{-[2pr^3 + 2p^2r^2(1-p) + 2p^2r^3(1-p) + 4p^3r^3(1-p) + 8(1-p)r^2pq + 2p^3r^2]\mu(000?)}\\
&\underbrace{-[2p^2r^2(1-p) + 2p^3r^3(1-p) + 4(1-p)pr^2q + 2p^3r^2]\mu(0000?) - 2p^3r^2\mu(00000?)}\\
&\underbrace{+[2p^2r^2(1-p) + 2p^2r^3(1-p) + 4p^3r^3(1-p) - 4(1-p)r^2q(r-p) + 2p^3r^2]\mu(000?1)}\\
&\underbrace{+[4p^3r^3(1-p) + 2p^3r^2]\mu(000?\{0,?\}1) + [2p^3r^3(1-p) + 2p^3r^2]\mu(0000?1)}\\
&\underbrace{-2pqr^2\mu(\widehat{***}) - 2prC_{10?} - 2pC_{100?}},
\end{aligned} \tag{4.38}$$

where we let B_2 denote the sum of the terms of (4.38) that have been grouped using underbraces, and we let B_1 denote the remaining terms. In other words,

$$\begin{aligned}
B_1 = w_2(\mu) &- [p(1-r)+q]\mu(10?) - [2q + r(4q^2 + 4qp^2 - 2q^3 + 2p^2 + 2q^2p)]\mu(100?)\\
&- 2r(1-p^2)\mu(1??1) - r(1-p)\mu(1?1) - r(4q^2 + 4qp^2 - 2q^3 + 2q^2p)[\mu(1???) + \mu(1??0)\\
&+ \mu(1?0?) + \mu(10??)] - r(4q^2 + 4qp^2 - 2q^3 + 2p^2 + 2q^2p)[\mu(1?00) + \mu(10?0)] - q\mu(?)\\
&- [p(1-r)+q][\mu(0?) + \mu(00?)] - pr^2(1+2q)\mu(1\widehat{***}1),
\end{aligned} \tag{4.39}$$

and

$$\begin{aligned}
B_2 = \ &\underbrace{2qp^2r\mu(\widehat{***})} - pr^2(1+2q)\mu(?000?) - 2qpr^2\mu(000\widehat{**}) \underbrace{+2pr^2(1+p+2q)\mu(000?)}\\
&- 2pr^2(1+2q+p)\mu(1?000) \underbrace{+2p^2r^2\mu(0000?) + 2p^2r^2\mu(00000?)} - 2p^2r^2\mu(000\widehat{**}1)\\
&- [2pr^3 + 2p^2r^2(1-p) + 2p^2r^3(1-p) + 4p^3r^3(1-p) + 8(1-p)r^2pq + 2p^3r^2]\mu(000?)\\
&- [2p^2r^2(1-p) + 2p^3r^3(1-p) + 4(1-p)pr^2q + 2p^3r^2]\mu(0000?) - 2p^3r^2\mu(00000?)\\
&\underbrace{+[2p^2r^2(1-p) + 2p^2r^3(1-p) + 4p^3r^3(1-p) - 4(1-p)r^2q(r-p) + 2p^3r^2]\mu(000?1)}\\
&\underbrace{+[4p^3r^3(1-p) + 2p^3r^2]\mu(000?\{0,?\}1) + [2p^3r^3(1-p) + 2p^3r^2]\mu(0000?1)}\\
&- 2pqr^2\mu(\widehat{***}) - 2prC_{10?} - 2pC_{100?}.
\end{aligned} \tag{4.40}$$

It is worthwhile to note here that all of the terms, other than $w_2(\mu)$, in B_1 are non-positive, whereas in B_2, the possibly non-negative (we write 'possibly' because the coefficient of $\mu(000?1)$ may or may not be non-negative, depending on the values of p and q) terms have been highlighted using underbraces in (4.40). We now have to make sure that these non-negative terms are negated using the existing non-positive terms on the right side of (4.40), and this is what we accomplish, to some extent, in §4.5.4.

4.5.4 Step 4 of analyzing (4.25)

We dedicate §4.5.4 to the analysis of B_2 in (4.40). Before we embark on this task, we perform a couple of rather intricate algebraic simplifications that are going to be of use while analysing B_2. The first of these is as follows, and this will be used in combining the coefficients of the various terms involving $\mu(000?)$ in the analysis of B_2:

$$\begin{aligned}
&- 2qpr^2 + 2pr^2(1+p+2q) - [2pr^3 + 2p^2r^2(1-p) + 2p^2r^3(1-p) + 4p^3r^3(1-p)\\
&+ 8(1-p)r^2pq + 2p^3r^2]\\
= &- 2qpr^2 + 2pr^2 + 2p^2r^2 + 4pqr^2 - 2pr^3 - 2p^2r^2(1-p) - 2p^2r^3(1-p) - 4p^3r^3(1-p)\\
&- 8(1-p)r^2pq - 2p^3r^2\\
= &\ 2pr^2 - 2pr^3 + 2p^2r^2 - 2p^2r^2(1-p) - 2p^3r^2 - 2p^2r^3(1-p) - 4p^3r^3(1-p)\\
&- 8(1-p)r^2pq + 2qpr^2
\end{aligned}$$

$$
\begin{aligned}
&= 2pr^2(1-r) \underbrace{+2p^2r^2 - 2p^2r^2}_{\text{cancel each other}} \underbrace{+2p^3r^2 - 2p^3r^2}_{\text{cancel each other}} -2p^2r^3(1-p) - 4p^3r^3(1-p) - 8pqr^2 \\
&\quad + 8p^2qr^2 + 2qpr^2 \\
&= 2p^2r^2 + 2pqr^2 - 2p^2r^3 + 2p^3r^3 - 4p^3r^3 + 4p^4r^3 - 6pqr^2 + 8p^2qr^2 \\
&= 2p^2r^2 - 2p^2r^3 - 2p^3r^3 + 4p^4r^3 - 4pqr^2 + 8p^2qr^2 \\
&= 2p^2r^2(p+q) - 2p^3r^3 + 4p^4r^3 - 4pqr^2 + 8p^2qr^2 \\
&= 2p^3r^2 + 2p^2qr^2 - 2p^3r^3 + 4p^4r^3 - 4pqr^2 + 8p^2qr^2 \\
&= 2p^3r^2 - 2p^3r^3 + 4p^4r^3 - 4pqr^2 + 10p^2qr^2 \\
&= 2p^3r^2(p+q) + 4p^4r^3 - 4pqr^2 + 10p^2qr^2 \\
&= \underbrace{2p^4r^2}_{(1)} \underbrace{+2p^3r^2q}_{(2)} +4p^4r^3 \underbrace{-4pqr^2 + 10p^2qr^2}_{(3)} \\
&= \underbrace{2pqr^2[p^2 - 2 + 5p] + 6p^4r^2 - 4p^4r^2}_{\text{combining (1), (2) and (3)}} +4p^4r^3 \\
&= 2pqr^2[p^2 - 2 + 5p] + 6p^4r^2 - 4p^4r^2(1-r) \\
&= 2pqr^2[p^2 - 2 + 5p] + 6p^4r^2 - 4p^4r^2(p+q).
\end{aligned} \tag{4.41}
$$

The second algebraic simplification we detail here is as follows, and this will be utilized in combining the coefficients of the various terms involving $\mu(000?1)$ in the analysis of B_2:

$$
\begin{aligned}
&2qpr^2 - 2pr^2(1+2q+p) + [2p^2r^2(1-p) + 2p^2r^3(1-p) + 4p^3r^3(1-p) \\
&\quad - 4(1-p)r^2q(r-p) + 2p^3r^2] \\
&= 2pr^2[q - 1 - 2q - p + p(1-p) + pr(1-p) + 2p^2r(1-p) + p^2] - 4(1-p)r^2q(r-p) \\
&= -2pr^2[1 + q + p - p + p^2 - p(1-p-q)(1-p) - 2p^2r(1-p) - p^2] - 4(1-p)r^2q(r-p) \\
&= -2pr^2[1 + q - p(1-p)^2 + pq(1-p) - 2p^2r(1-p)] - 4(1-p)r^2q(r-p) \\
&= -2pr^2[1 + q - p(1-2p+p^2) + pq(1-p) - 2p^2r(1-p)] - 4(1-p)r^2q(r-p) \\
&= -2pr^2[1 + q - p + 2p^2 - p^3 + pq(1-p) - 2p^2r(1-p)] - 4(1-p)r^2q(r-p) \\
&= -2pr^2[1 + q - p + 2p^2 - 2p^2r(1-p) + pq(1-p) - p^3] - 4(1-p)r^2q(r-p).
\end{aligned} \tag{4.42}
$$

In the analysis that follow, in many steps, we indicate using underbraces the terms that are to be combined in the next step, and we use the following identities:

1. $\mu(000\widehat{**}) = \mu(000?) - \mu(000?1) + \mu(0000?)$,
2. $\mu(0000?) = \mu(00000?) + \mu(?0000?) + \mu(10000?)$,
3. $\mu(000\widehat{**}1) = \mu(000?\{0,?\}1) + \mu(0000?1)$.

The simplification of B_2 continues as follows:

$$
\begin{aligned}
B_2 &= 2qp^2r\mu(\widehat{***}) - pr^2(1+2q)\mu(?000?) \underbrace{-2qpr^2\mu(000\widehat{**})}_{\text{use (1)}} \underbrace{+2pr^2(1+p+2q)\mu(000?)}_{\text{term involving } \mu(000?)} \\
&\quad - 2pr^2(1+2q+p)\mu(1?000) + 2p^2r^2\mu(0000?) + 2p^2r^2\mu(00000?) - 2p^2r^2\mu(000\widehat{**}1) \\
&\quad \underbrace{-[2pr^3 + 2p^2r^2(1-p) + 2p^2r^3(1-p) + 4p^3r^3(1-p) + 8(1-p)r^2pq + 2p^3r^2]\mu(000?)}_{\text{term involving } \mu(000?)} \\
&\quad - [2p^2r^2(1-p) + 2p^3r^3(1-p) + 4(1-p)pr^2q + 2p^3r^2]\mu(0000?) - 2p^3r^2\mu(00000?) \\
&\quad + [2p^2r^2(1-p) + 2p^2r^3(1-p) + 4p^3r^3(1-p) - 4(1-p)r^2q(r-p) + 2p^3r^2]\mu(000?1) \\
&\quad + [4p^3r^3(1-p) + 2p^3r^2]\mu(000?\{0,?\}1) + [2p^3r^3(1-p) + 2p^3r^2]\mu(0000?1)
\end{aligned}
$$

$$
\begin{aligned}
& - 2pqr^2\mu(\widehat{***}) - 2prC_{10?} - 2pC_{100?} \\
= {} & 2qp^2r\mu(\widehat{***}) - pr^2(1+2q)\mu(?000?) \underbrace{-2qpr^2\mu(0000?) + 2qpr^2\mu(000?1)}_{\text{obtained from (1)}} \\
& - 2pr^2(1+2q+p)\mu(1?000) + 2p^2r^2\mu(0000?) + 2p^2r^2\mu(00000?) - 2p^2r^2\mu(000\widehat{**}1) \\
& \qquad \underbrace{+[2pqr^2\{5p-2+p^2\} + 6p^4r^2 - 4p^4r^2(p+q)]\mu(000?)}_{\text{summing terms involving } \mu(000?) \text{, including the one obtained from (1), and using (4.41)}} \\
& - [2p^2r^2(1-p) + 2p^3r^3(1-p) + 4(1-p)pr^2q + 2p^3r^2]\mu(0000?) - 2p^3r^2\mu(00000?) \\
& + [2p^2r^2(1-p) + 2p^2r^3(1-p) + 4p^3r^3(1-p) - 4(1-p)r^2q(r-p) + 2p^3r^2]\mu(000?1) \\
& + [4p^3r^3(1-p) + 2p^3r^2]\mu(000?\{0,?\}1) + [2p^3r^3(1-p) + 2p^3r^2]\mu(0000?1) \\
& - 2pqr^2\mu(\widehat{***}) - 2prC_{10?} - 2pC_{100?} \\
= {} & 2qp^2r\mu(\widehat{***}) - pr^2(1+2q)\mu(?000?) \underbrace{-2qpr^2\mu(0000?)}_{\text{term involving } \mu(0000?)} +2qpr^2\mu(000?1) \\
& - 2pr^2(1+2q+p)\mu(1?000) \underbrace{+2p^2r^2\mu(0000?)}_{\text{term involving } \mu(0000?)} +2p^2r^2\mu(00000?) - 2p^2r^2\mu(000\widehat{**}1) \\
& + [2pqr^2\{5p-2+p^2\} + 6p^4r^2 - 4p^4r^2(p+q)]\mu(000?) \\
& \underbrace{-[2p^2r^2(1-p) + 2p^3r^3(1-p) + 4(1-p)pr^2q + 2p^3r^2]\mu(0000?)}_{\text{term involving } \mu(0000?)} -2p^3r^2\mu(00000?) \\
& + [2p^2r^2(1-p) + 2p^2r^3(1-p) + 4p^3r^3(1-p) - 4(1-p)r^2q(r-p) + 2p^3r^2]\mu(000?1) \\
& + [4p^3r^3(1-p) + 2p^3r^2]\mu(000?\{0,?\}1) + [2p^3r^3(1-p) + 2p^3r^2]\mu(0000?1) \\
& - 2pqr^2\mu(\widehat{***}) - 2prC_{10?} - 2pC_{100?} \\
= {} & 2qp^2r\mu(\widehat{***}) - pr^2(1+2q)\mu(?000?) + 2qpr^2\mu(000?1) - 2pr^2(1+2q+p)\mu(1?000) \\
& + 2p^2r^2\mu(00000?) - 2p^2r^2\mu(000\widehat{**}1) \\
& + [2pqr^2\{5p-2+p^2\} + 6p^4r^2 - 4p^4r^2(p+q)]\mu(000?) \\
& \underbrace{-(6pqr^2 - 4p^2qr^2)\mu(0000?) - 2p^3r^3(1-p)\mu(0000?)}_{\text{summing terms involving } \mu(0000?)} -2p^3r^2\mu(00000?) \\
& + [2p^2r^2(1-p) + 2p^2r^3(1-p) + 4p^3r^3(1-p) - 4(1-p)r^2q(r-p) + 2p^3r^2]\mu(000?1) \\
& + [4p^3r^3(1-p) + 2p^3r^2]\mu(000?\{0,?\}1) + [2p^3r^3(1-p) + 2p^3r^2]\mu(0000?1) \\
& - 2pqr^2\mu(\widehat{***}) - 2prC_{10?} - 2pC_{100?} \\
= {} & 2qp^2r\mu(\widehat{***}) - pr^2(1+2q)\mu(?000?) + 2qpr^2\mu(000?1) - 2pr^2(1+2q+p)\mu(1?000) \\
& \underbrace{+2p^2r^2\mu(00000?)}_{\text{term involving } \mu(00000?)} -2p^2r^2\mu(000\widehat{**}1) \\
& + [2pqr^2\{5p-2+p^2\} + 6p^4r^2 - 4p^4r^2(p+q)]\mu(000?) \\
& - (6pqr^2 - 4p^2qr^2)\mu(0000?) \underbrace{-2p^3r^3(1-p)\mu(0000?)}_{\text{use (2)}} \underbrace{-2p^3r^2\mu(00000?)}_{\text{term involving } \mu(00000?)} \\
& + [2p^2r^2(1-p) + 2p^2r^3(1-p) + 4p^3r^3(1-p) - 4(1-p)r^2q(r-p) + 2p^3r^2]\mu(000?1) \\
& + [4p^3r^3(1-p) + 2p^3r^2]\mu(000?\{0,?\}1) + [2p^3r^3(1-p) + 2p^3r^2]\mu(0000?1) \\
& - 2pqr^2\mu(\widehat{***}) - 2prC_{10?} - 2pC_{100?} \\
= {} & 2qp^2r\mu(\widehat{***}) - pr^2(1+2q)\mu(?000?) + 2qpr^2\mu(000?1) - 2pr^2(1+2q+p)\mu(1?000) \\
& - 2p^2r^2\mu(000\widehat{**}1) + [2pqr^2\{5p-2+p^2\} + 6p^4r^2 - 4p^4r^2(p+q)]\mu(000?) \\
& - (6pqr^2 - 4p^2qr^2)\mu(0000?) \qquad \underbrace{+2p^2r^2(1-p)(1-pr)\mu(00000?)}_{\text{summing terms involving } \mu(00000?) \text{, including that obtained from (2)}}
\end{aligned}
$$

$$
\begin{aligned}
&\underbrace{-2p^3r^3(1-p)\{\mu(10000?)+\mu(?0000?)\}}_{\text{remaining terms obtained by applying (2)}}\\
&+[2p^2r^2(1-p)+2p^2r^3(1-p)+4p^3r^3(1-p)-4(1-p)r^2q(r-p)+2p^3r^2]\mu(000?1)\\
&+[4p^3r^3(1-p)+2p^3r^2]\mu(000?\{0,?\}1)+[2p^3r^3(1-p)+2p^3r^2]\mu(0000?1)\\
&-2pqr^2\mu(\widehat{***})-2prC_{10?}-2pC_{100?}\\
=\ &2qp^2r\mu(\widehat{***})-pr^2(1+2q)\mu(?000?)\underbrace{+2qpr^2\mu(000?1)-2pr^2(1+2q+p)\mu(1?000)}_{\text{terms involving }\mu(000?1)}\\
&-2p^2r^2\mu(000\widehat{**}1)+[2pqr^2\{5p-2+p^2\}+6p^4r^2-4p^4r^2(p+q)]\mu(000?)\\
&-(6pqr^2-4p^2qr^2)\mu(0000?)+2p^2r^2(1-p)(1-pr)\mu(00000?)\\
&-2p^3r^3(1-p)\{\mu(10000?)+\mu(?0000?)\}\\
&\underbrace{+[2p^2r^2(1-p)+2p^2r^3(1-p)+4p^3r^3(1-p)-4(1-p)r^2q(r-p)+2p^3r^2]\mu(000?1)}_{\text{term involving }\mu(000?1)}\\
&+[4p^3r^3(1-p)+2p^3r^2]\mu(000?\{0,?\}1)+[2p^3r^3(1-p)+2p^3r^2]\mu(0000?1)\\
&-2pqr^2\mu(\widehat{***})-2prC_{10?}-2pC_{100?}\\
=\ &2qp^2r\mu(\widehat{***})-pr^2(1+2q)\mu(?000?)-2p^2r^2\underbrace{\mu(000\widehat{**}1)}_{\text{split by (3)}}+[2pqr^2\{5p-2+p^2\}+6p^4r^2\\
&-4p^4r^2(p+q)]\mu(000?)-(6pqr^2-4p^2qr^2)\mu(0000?)+2p^2r^2(1-p)(1-pr)\mu(00000?)\\
&-2p^3r^3(1-p)\{\mu(10000?)+\mu(?0000?)\}\\
&\underbrace{-2pr^2\{1+q-p+2p^2-2p^2r(1-p)+pq(1-p)-p^3\}\mu(000?1)-4(1-p)r^2q(r-p)\mu(000?1)}_{\text{summing terms involving }\mu(000?1)\text{ and using (4.42)}}\\
&+[4p^3r^3(1-p)+2p^3r^2]\mu(000?\{0,?\}1)+[2p^3r^3(1-p)+2p^3r^2]\mu(0000?1)\\
&-2pqr^2\mu(\widehat{***})-2prC_{10?}-2pC_{100?}\\
=\ &2qp^2r\mu(\widehat{***})-pr^2(1+2q)\mu(?000?)\underbrace{-2p^2r^2\mu(000?\{0,?\}1)-2p^2r^2\mu(0000?1)}_{\text{to be added to underbraced terms below}}\\
&+[2pqr^2\{5p-2+p^2\}+6p^4r^2-4p^4r^2(p+q)]\mu(000?)\\
&-(6pqr^2-4p^2qr^2)\mu(0000?)+2p^2r^2(1-p)(1-pr)\mu(00000?)\\
&-2p^3r^3(1-p)\{\mu(10000?)+\mu(?0000?)\}\\
&-2pr^2\{1+q-p+2p^2-2p^2r(1-p)+pq(1-p)-p^3\}\mu(000?1)\\
&-4(1-p)r^2q(r-p)\mu(000?1)\\
&\underbrace{+[4p^3r^3(1-p)+2p^3r^2]\mu(000?\{0,?\}1)+[2p^3r^3(1-p)+2p^3r^2]\mu(0000?1)}_{\text{to be added to underbraced terms above}}\\
&-2pqr^2\mu(\widehat{***})-2prC_{10?}-2pC_{100?}\\
=\ &2qp^2r\mu(\widehat{***})-pr^2(1+2q)\mu(?000?)\\
&+[2pqr^2\{5p-2+p^2\}+6p^4r^2-4p^4r^2(p+q)]\mu(000?)\\
&-(6pqr^2-4p^2qr^2)\mu(0000?)+2p^2r^2(1-p)(1-pr)\mu(00000?)\\
&-2p^3r^3(1-p)\{\mu(10000?)+\mu(?0000?)\}\\
&-2pr^2\{1+q-p+2p^2-2p^2r(1-p)+pq(1-p)-p^3\}\mu(000?1)\\
&-4(1-p)r^2q(r-p)\mu(000?1)\\
&\underbrace{-2p^2r^2(1-p)(1-2pr)\mu(000?\{0,?\}1)-2p^2r^2(1-p)(1-pr)\mu(0000?1)}_{\text{summing terms highlighted by underbraces in the previous step}}\\
&-2pqr^2\mu(\widehat{***})-2prC_{10?}-2pC_{100?}. \qquad (4.43)
\end{aligned}
$$

We now pause for a bit to write down explicitly $-2prC_{10?}-2pC_{100?}$, where we borrow the mathematical expressions for $C_{10?}$ and $C_{100?}$ from (4.30) and (4.36) derived in §4.5.2:

$$\begin{aligned}
-2prC_{10?}-2pC_{100?} = &-2qp^2r^2\mu(1\{0,?\}\widehat{***})-2pq(1-q)r^2\mu(1\widehat{***})-2qp^3r\mu(\widehat{***})\\
&-2qp^2r^2\mu(1\{0,?\}\widehat{***})-2pqr^2(1-q+p)\mu(1\widehat{***})\\
= &-4qp^2r^2\mu(1\{0,?\}\widehat{***})-2pqr^2(2-2q+p)\mu(1\widehat{***})-2qp^3r\mu(\widehat{***}).
\end{aligned} \tag{4.44}$$

Substituting (4.44) in (4.43), we obtain:

$$\begin{aligned}
B_2 = &\underbrace{2qp^2r\mu(\widehat{***})}_{\text{term involving } \mu(\widehat{***})} -pr^2(1+2q)\mu(?000?)+[2pqr^2\{5p-2+p^2\}+6p^4r^2\\
&-4p^4r^2(p+q)]\mu(000?)-(6pqr^2-4p^2qr^2)\mu(0000?)+2p^2r^2(1-p)(1-pr)\mu(00000?)\\
&-2p^3r^3(1-p)\{\mu(10000?)+\mu(?0000?)\}\\
&-2pr^2\{1+q-p+2p^2-2p^2r(1-p)+pq(1-p)-p^3\}\mu(000?1)\\
&-4(1-p)r^2q(r-p)\mu(000?1)-2p^2r^2(1-p)(1-2pr)\mu(000?\{0,?\}1)\\
&-2p^2r^2(1-p)(1-pr)\mu(0000?1) \underbrace{-2pqr^2\mu(\widehat{***})}_{\text{term involving } \mu(\widehat{***})}\\
&-4qp^2r^2\mu(1\{0,?\}\widehat{***})-2pqr^2(2-2q+p)\mu(1\widehat{***}) \underbrace{-2qp^3r\mu(\widehat{***})}_{\text{term involving } \mu(\widehat{***})}\\
= &\underbrace{2pqr(p-r-p^2)\mu(\widehat{***})}_{\text{adding terms involving } \mu(\widehat{***})} -pr^2(1+2q)\mu(?000?)\\
&+[2pqr^2\{5p-2+p^2\}+6p^4r^2-4p^4r^2(p+q)]\mu(000?)-(6pqr^2-4p^2qr^2)\mu(0000?)\\
&+2p^2r^2(1-p)(1-pr)\mu(00000?)-2p^3r^3(1-p)\{\mu(10000?)+\mu(?0000?)\}\\
&-2pr^2\{1+q-p+2p^2-2p^2r(1-p)+pq(1-p)-p^3\}\mu(000?1)\\
&-4(1-p)r^2q(r-p)\mu(000?1)\\
&-2p^2r^2(1-p)(1-2pr)\mu(000?\{0,?\}1)-2p^2r^2(1-p)(1-pr)\mu(0000?1)\\
&-4qp^2r^2 \underbrace{\mu(1\{0,?\}\widehat{***})}_{\text{split into } \mu(1000?) \text{ and the rest}} -2pqr^2(2-2q+p)\mu(1\widehat{***})\\
= &\,2pqr(p-r-p^2)\mu(\widehat{***})\underbrace{-pr^2(1+2q)\mu(?000?)}\\
&+[2pqr^2\{5p-2+p^2\}+6p^4r^2-4p^4r^2(p+q)]\mu(000?)\underbrace{-(6pqr^2-4p^2qr^2)\mu(0000?)}\\
&+2p^2r^2(1-p)(1-pr)\mu(00000?)-2p^3r^3(1-p)\{\mu(10000?)+\mu(?0000?)\}\\
&-2pr^2\{1+q-p+2p^2-2p^2r(1-p)+pq(1-p)-p^3\}\mu(000?1)\\
&-4(1-p)r^2q(r-p)\mu(000?1)-2p^2r^2(1-p)(1-2pr)\mu(000?\{0,?\}1)\\
&-2p^2r^2(1-p)(1-pr)\mu(0000?1)\underbrace{-4qp^2r^2\mu(1000?)}\\
&-4qp^2r^2[\mu(1\{0,?\}\widehat{***})-\mu(1000?)]-2pqr^2(2-2q+p)\mu(1\widehat{***}).
\end{aligned} \tag{4.45}$$

We show here how we combine the terms highlighted by underbraces in the last step of (4.45):

$$\begin{aligned}
&-pr^2(1+2q)\mu(?000?)-(6pqr^2-4p^2qr^2)\mu(0000?)-4qp^2r^2\mu(1000?)\\
= &\underbrace{-2p^2qr^2\mu(?000?)-pr^2\{1+2q(1-p)\}\mu(?000?)}_{\text{splitting the term involving } \mu(?000?)}\underbrace{-2p^2qr^2\mu(0000?)-6pqr^2(1-p)\mu(0000?)}_{\text{splitting the term involving } \mu(0000?)}\\
&-4p^2qr^2\mu(1000?)
\end{aligned}$$

$$= \underbrace{-2p^2qr^2\mu(?000?) - 2p^2qr^2\mu(0000?) - 2p^2qr^2\mu(1000?)} - pr^2\{1+2q(1-p)\}\mu(?000?)$$
$$- 6pqr^2(1-p)\mu(0000?) - 2p^2qr^2\mu(1000?)$$
$$= -2p^2qr^2\mu(000?) - pr^2\{1+2q(1-p)\}\mu(?000?) - 6pqr^2(1-p)\mu(0000?) - 2p^2qr^2\mu(1000?). \tag{4.46}$$

Substituting (4.46) in (4.45) yields (we have highlighted the terms that have come from (4.46) using underbraces in the very first step of the computation below):

$$\begin{aligned}
B_2 &= 2pqr(p-r-p^2)\mu(\widehat{***}) \underbrace{-2p^2qr^2\mu(000?) - pr^2\{1+2q(1-p)\}\mu(?000?)} \\
&\quad \underbrace{-6pqr^2(1-p)\mu(0000?)} + [2pqr^2\{5p-2+p^2\} + 6p^4r^2 - 4p^4r^2(p+q)]\mu(000?) \\
&\quad + 2p^2r^2(1-p)(1-pr)\mu(00000?) - 2p^3r^3(1-p)\{\mu(10000?) + \mu(?0000?)\} \\
&\quad - 2pr^2\{1+q-p+2p^2-2p^2r(1-p)+pq(1-p)-p^3\}\mu(000?1) \\
&\quad - 4(1-p)r^2q(r-p)\mu(000?1) - 2p^2r^2(1-p)(1-2pr)\mu(000?\{0,?\}1) \\
&\quad - 2p^2r^2(1-p)(1-pr)\mu(0000?1) \underbrace{-2p^2qr^2\mu(1000?)} \\
&\quad - 4qp^2r^2[\mu(1\{0,?\}\widehat{***}) - \mu(1000?)] - 2pqr^2(2-2q+p)\mu(1\widehat{***}) \\
&= 2pqr(p-r-p^2)\mu(\widehat{***}) \underbrace{-2p^2qr^2\mu(000?)} - pr^2\{1+2q(1-p)\}\mu(?000?) \\
&\quad - 6pqr^2(1-p)\mu(0000?) \underbrace{+[2pqr^2\{5p-2+p^2\} + 6p^4r^2 - 4p^4r^2(p+q)]\mu(000?)} \\
&\quad + 2p^2r^2(1-p)(1-pr)\mu(00000?) - 2p^3r^3(1-p)\{\mu(10000?) + \mu(?0000?)\} \\
&\quad - 2pr^2\{1+q-p+2p^2-2p^2r(1-p)+pq(1-p)-p^3\}\mu(000?1) \\
&\quad - 4(1-p)r^2q(r-p)\mu(000?1) - 2p^2r^2(1-p)(1-2pr)\mu(000?\{0,?\}1) \\
&\quad - 2p^2r^2(1-p)(1-pr)\mu(0000?1) - 2p^2qr^2\mu(1000?) \\
&\quad - 4qp^2r^2[\mu(1\{0,?\}\widehat{***}) - \mu(1000?)] - 2pqr^2(2-2q+p)\mu(1\widehat{***}) \\
&= 2pqr(p-r-p^2)\mu(\widehat{***}) - pr^2\{1+2q(1-p)\}\mu(?000?) - 6pqr^2(1-p)\mu(0000?) \\
&\quad \underbrace{+[2pqr^2\{4p-2+p^2\} + 6p^4r^2 - 4p^4r^2(p+q)]\mu(000?)}_{\text{summing terms involving } \mu(000?)} \\
&\quad + 2p^2r^2(1-p)(1-pr)\mu(00000?) - 2p^3r^3(1-p)\{\mu(10000?) + \mu(?0000?)\} \\
&\quad - 2pr^2\{1+q-p+2p^2-2p^2r(1-p)+pq(1-p)-p^3\}\mu(000?1) \\
&\quad - 4(1-p)r^2q(r-p)\mu(000?1) - 2p^2r^2(1-p)(1-2pr)\mu(000?\{0,?\}1) \\
&\quad - 2p^2r^2(1-p)(1-pr)\mu(0000?1) - 2p^2qr^2\mu(1000?) \\
&\quad - 4qp^2r^2[\mu(1\{0,?\}\widehat{***}) - \mu(1000?)] - 2pqr^2(2-2q+p)\mu(1\widehat{***}).
\end{aligned} \tag{4.47}$$

The above, i.e. the last step of (4.47), serves as our final, simplified expression for B_2. From (4.38), (4.39) and (4.47), we see that our weight function inequality has now transformed into:

$$\begin{aligned}
&w_2(\widehat{\mathrm{F}}_{p,q}\,\mu) \leqslant w_2(\mu) - [p(1-r)+q]\mu(10?) - [2q + r(4q^2+4qp^2-2q^3+2p^2+2q^2p)]\mu(100?) \\
&- 2r(1-p^2)\mu(1??1) - r(1-p)\mu(1?1) - r(4q^2+4qp^2-2q^3+2q^2p)[\mu(1???) + \mu(1??0) \\
&+ \mu(1?0?) + \mu(10??)] - r(4q^2+4qp^2-2q^3+2p^2+2q^2p)[\mu(1?00) + \mu(10?0)] - q\mu(?) \\
&- [p(1-r)+q][\mu(0?) + \mu(00?)] - pr^2(1+2q)\mu(1\widehat{***}1) \\
&\underbrace{+2pqr(p-r-p^2)\mu(\widehat{***})} - pr^2\{1+2q(1-p)\}\mu(?000?) - 6pqr^2(1-p)\mu(0000?) \\
&\underbrace{+[2pqr^2\{4p-2+p^2\} + 6p^4r^2 - 4p^4r^2(p+q)]\mu(000?)} \\
&\underbrace{+2p^2r^2(1-p)(1-pr)\mu(00000?)} - 2p^3r^3(1-p)\{\mu(10000?) + \mu(?0000?)\}
\end{aligned}$$

$$
\begin{aligned}
&- 2pr^2\{1+q-p+2p^2-2p^2r(1-p)+pq(1-p)-p^3\}\mu(000?1)\\
&- 4(1-p)r^2q(r-p)\mu(000?1) - 2p^2r^2(1-p)(1-2pr)\mu(000?\{0,?\}1)\\
&- 2p^2r^2(1-p)(1-pr)\mu(0000?1) - 2p^2qr^2\mu(1000?)\\
&- 4qp^2r^2[\mu(1\{0,?\}\widehat{***}) - \mu(1000?)] - 2pqr^2(2-2q+p)\mu(1\widehat{***}). \qquad (4.48)
\end{aligned}
$$

The underbraces in the right side of (4.48) are intended to highlight the (possibly, depending on the values of p and q) non-negative terms that still remain. It is worthwhile to note, comparing (4.48) with (4.40), that we have already taken care of the (previously non-negative) terms involving $\mu(0000?)$, $\mu(000?1)$, $\mu(000?\{0,?\}1)$ and $\mu(0000?1)$, via the long and tedious algebraic manipulations that ultimately lead to (4.47).

4.6 The fourth step of composing the weight function

The third adjustment is carried out as follows:

$$
w_3(\mu) = w_2(\mu) - 2(q+p^2r)\mu(100?) - 2p^2r\{\mu(1?00)+\mu(10?0)\} - q\mu(?). \qquad (4.49)
$$

Writing (4.48) as $w_2(\widehat{\mathrm{F}}_{p,q}\,\mu) \leqslant w_2(\mu) + A_2$, implementing the adjustment described in (4.49), and using the idea presented in (4.16), we obtain the weight function inequality

$$
\begin{aligned}
w_3(\widehat{\mathrm{F}}_{p,q}\,\mu) \leqslant\ & w_3(\mu) + 2(q+p^2r)\mu(100?) + 2p^2r\{\mu(1?00)+\mu(10?0)\} + q\mu(?)\\
& - 2(q+p^2r)\,\widehat{\mathrm{F}}_{p,q}\,\mu(100?) - 2p^2r\{\widehat{\mathrm{F}}_{p,q}\,\mu(1?00) + \widehat{\mathrm{F}}_{p,q}\,\mu(10?0)\}\\
& - q\,\widehat{\mathrm{F}}_{p,q}\,\mu(?) + A_2. \qquad (4.50)
\end{aligned}
$$

As noted using underbraces in (4.48), we have to somehow negate the non-negative terms involving $\mu(000?)$, $\mu(00000?)$ and $\mu(\widehat{***})$ using existing non-positive terms on the right side of (4.48).

1. As seen from (4.52), one of the terms in the expansion of $\widehat{\mathrm{F}}_{p,q}\,\mu(1?00)$ involves $\mu(000?)$;
2. as seen from (4.53), one of the terms in the expansion of $\widehat{\mathrm{F}}_{p,q}\,\mu(10?0)$ involves $\mu(000\widehat{**})$, which can be written as $\mu(000?) + \mu(0000?) - \mu(000?1)$;
3. and finally, one of the terms in the expansion of $\widehat{\mathrm{F}}_{p,q}\,\mu(100?)$, as seen in (4.6), involves $\mu(000\widehat{***})$, which can be written as $\mu(000\widehat{***}) = \mu(000?) + \mu(0000?) + \mu(00000?) - \mu(000?1) - \mu(000\widehat{**}1)$.

Therefore, from (4.50), it can be hoped that $-2(q+p^2r)\,\widehat{\mathrm{F}}_{p,q}\,\mu(100?) - 2p^2r\{\widehat{\mathrm{F}}_{p,q}\,\mu(1?00) + \widehat{\mathrm{F}}_{p,q}\,\mu(10?0)\}$ will aid in negating the non-negative terms involving $\mu(000?)$ and $\mu(0000?)$ highlighted in (4.48) using underbraces. From (4.4), we see that $\widehat{\mathrm{F}}_{p,q}\,\mu(?) = r\mu(\widehat{***})$, so that $-q\,\widehat{\mathrm{F}}_{p,q}\,\mu(?)$ may aid in negating the non-negative term involving $\mu(\widehat{***})$ highlighted in (4.48) using underbraces.

At the same time, we have to make sure that we can afford to introduce the non-negative terms $2(q+p^2r)\mu(100?) + 2p^2r\{\mu(1?00)+\mu(10?0)\} + q\mu(?)$ in (4.50) without turning any of the existing non-positive terms in the right side of (4.48) non-negative:

1. the term $-[2q + r(4q^2+4qp^2-2q^3+2p^2+2q^2p)]\mu(100?)$ in the right side of (4.48) provides us with $-2(q+p^2r)\mu(100?)$,
2. whereas the term $-r(4q^2+4qp^2-2q^3+2p^2+2q^2p)[\mu(1?00)+\mu(10?0)]$ in the right side of (4.48) supplies us with $-2p^2r\{\mu(1?00)+\mu(10?0)\}$;
3. finally, the existing $-q\mu(?)$ term in the right side of (4.48) cancels out the $+q\mu(?)$ term.

We hope that the above two paragraphs serve to motivate the choice of our adjustment described in (4.49).

4.6.1 Step 1 of analyzing the effect of the adjustment in (4.49)

Incorporating (4.49) into (4.48), as illustrated in (4.50), we obtain (in any given step, we highlight, using underbraces, terms that are to be combined / manipulated algebraically to obtain the next step):

$$
\begin{aligned}
& w_3(\widehat{\mathrm{F}}_{p,q}\,\mu) \leqslant w_3(\mu) + 2(q+p^2r)\mu(100?) + 2p^2r\{\mu(1?00)+\mu(10?0)\} + q\mu(?) \\
& - [p(1-r)+q]\mu(10?) - [2q + r(4q^2+4qp^2-2q^3+2p^2+2q^2p)]\mu(100?) \\
& - 2r(1-p^2)\mu(1??1) \\
& - r(1-p)\mu(1?1) - r(4q^2+4qp^2-2q^3+2q^2p)[\mu(1???)+\mu(1??0)+\mu(1?0?)+\mu(10??)] \\
& - r(4q^2+4qp^2-2q^3+2p^2+2q^2p)[\mu(1?00)+\mu(10?0)] - q\mu(?) \\
& - [p(1-r)+q][\mu(0?)+\mu(00?)] - pr^2(1+2q)\mu(1\widehat{***}1) + 2pqr(p-r-p^2)\mu(\widehat{***}) \\
& - pr^2\{1+2q(1-p)\}\mu(?000?) - 6pqr^2(1-p)\mu(0000?) \\
& + [2pqr^2\{4p-2+p^2\} + 6p^4r^2 - 4p^4r^2(p+q)]\mu(000?) \\
& + 2p^2r^2(1-p)(1-pr)\mu(00000?) - 2p^3r^3(1-p)\{\mu(10000?)+\mu(?0000?)\} \\
& - 2pr^2\{1+q-p+2p^2-2p^2r(1-p)+pq(1-p)-p^3\}\mu(000?1) \\
& - 4(1-p)r^2q(r-p)\mu(000?1) - 2p^2r^2(1-p)(1-2pr)\mu(000?\{0,?\}1) \\
& - 2p^2r^2(1-p)(1-pr)\mu(0000?1) - 2p^2qr^2\mu(1000?) \\
& - 4qp^2r^2[\mu(1\{0,?\}\widehat{***}) - \mu(1000?)] - 2pqr^2(2-2q+p)\mu(1\widehat{***}) \\
& - 2(q+p^2r)\,\widehat{\mathrm{F}}_{p,q}\,\mu(100?) - 2p^2r\{\widehat{\mathrm{F}}_{p,q}\,\mu(1?00) + \widehat{\mathrm{F}}_{p,q}\,\mu(10?0)\} - q\,\widehat{\mathrm{F}}_{p,q}\,\mu(?) \\
= {} & w_3(\mu) - [p(1-r)+q]\mu(10?) \underbrace{+2(q+p^2r)\mu(100?) + 2p^2r\{\mu(1?00)+\mu(10?0)\} + q\mu(?)} \\
& \underbrace{-[2q + r(4q^2+4qp^2-2q^3+2p^2+2q^2p)]\mu(100?)} -2r(1-p^2)\mu(1??1) - r(1-p)\mu(1?1) \\
& \underbrace{-r(4q^2+4qp^2-2q^3+2q^2p)[\mu(1???)+\mu(1??0)+\mu(1?0?)+\mu(10??)]} \\
& \underbrace{-r(4q^2+4qp^2-2q^3+2p^2+2q^2p)[\mu(1?00)+\mu(10?0)] - q\mu(?)} \\
& - [p(1-r)+q][\mu(0?)+\mu(00?)] - pr^2(1+2q)\mu(1\widehat{***}1) \\
& + 2pqr(p-r-p^2)\mu(\widehat{***}) - pr^2\{1+2q(1-p)\}\mu(?000?) - 6pqr^2(1-p)\mu(0000?) \\
& + [2pqr^2\{4p-2+p^2\} + 6p^4r^2 - 4p^4r^2(p+q)]\mu(000?) \\
& + 2p^2r^2(1-p)(1-pr)\mu(00000?) - 2p^3r^3(1-p)\{\mu(10000?)+\mu(?0000?)\} \\
& - 2pr^2\{1+q-p+2p^2-2p^2r(1-p)+pq(1-p)-p^3\}\mu(000?1) \\
& - 4(1-p)r^2q(r-p)\mu(000?1) - 2p^2r^2(1-p)(1-2pr)\mu(000?\{0,?\}1) \\
& - 2p^2r^2(1-p)(1-pr)\mu(0000?1) - 2p^2qr^2\mu(1000?) \\
& - 4qp^2r^2[\mu(1\{0,?\}\widehat{***}) - \mu(1000?)] - 2pqr^2(2-2q+p)\mu(1\widehat{***}) \\
& - 2(q+p^2r)\,\widehat{\mathrm{F}}_{p,q}\,\mu(100?) - 2p^2r\{\widehat{\mathrm{F}}_{p,q}\,\mu(1?00) + \widehat{\mathrm{F}}_{p,q}\,\mu(10?0)\} - q\,\widehat{\mathrm{F}}_{p,q}\,\mu(?) \\
= {} & w_3(\mu) - [p(1-r)+q]\mu(10?) \underbrace{-r(4q^2+4qp^2-2q^3+2q^2p)\mu(1\widehat{***})}_{\text{combining all terms indicated by underbraces above}} -2r(1-p^2)\mu(1??1) \\
& - r(1-p)\mu(1?1) - [p(1-r)+q][\mu(0?)+\mu(00?)] - pr^2(1+2q)\mu(1\widehat{***}1) \\
& + 2pqr(p-r-p^2)\mu(\widehat{***}) - pr^2\{1+2q(1-p)\}\mu(?000?) - 6pqr^2(1-p)\mu(0000?) \\
& + [2pqr^2\{4p-2+p^2\} + 6p^4r^2 - 4p^4r^2(p+q)]\mu(000?) \\
& + 2p^2r^2(1-p)(1-pr)\mu(00000?) - 2p^3r^3(1-p)\{\mu(10000?)+\mu(?0000?)\} \\
& - 2pr^2\{1+q-p+2p^2-2p^2r(1-p)+pq(1-p)-p^3\}\mu(000?1) \\
& - 4(1-p)r^2q(r-p)\mu(000?1) - 2p^2r^2(1-p)(1-2pr)\mu(000?\{0,?\}1)
\end{aligned}
$$

$$
\begin{aligned}
&- 2p^2r^2(1-p)(1-pr)\mu(0000?1) - 2p^2qr^2\mu(1000?)\\
&- 4qp^2r^2[\mu(1\{0,?\}\widehat{***}) - \mu(1000?)]\\
&- 2pqr^2(2-2q+p)\mu(1\widehat{***}) \underbrace{-2(q+p^2r)\,\widehat{\mathrm{F}}_{p,q}\,\mu(100?)}\\
&\underbrace{-2p^2r\{\widehat{\mathrm{F}}_{p,q}\,\mu(1?00) + \widehat{\mathrm{F}}_{p,q}\,\mu(10?0)\} - q\,\widehat{\mathrm{F}}_{p,q}\,\mu(?)}. \qquad (4.51)
\end{aligned}
$$

4.6.2 Step 2 of analyzing the effect of the adjustment in (4.49)

Our next task is to compute the sum of the *last three terms*, highlighted by underbraces, in the last step of (4.51). To this end, we need to compute $\widehat{\mathrm{F}}_{p,q}\,\mu(1?00)$ and $\widehat{\mathrm{F}}_{p,q}\,\mu(10?0)$. As was the case for the computations carried out in §4.5.2, we are only concerned with parts of the expressions for these probabilities. While computing $\widehat{\mathrm{F}}_{p,q}\,\mu(1?00)$, we consider only those cases in which $\eta(0)=\eta(1)=\eta(2)=0$ leads to the event that $(\widehat{\mathrm{F}}_{p,q}\,\eta(0), \widehat{\mathrm{F}}_{p,q}\,\eta(1), \widehat{\mathrm{F}}_{p,q}\,\eta(2), \widehat{\mathrm{F}}_{p,q}\,\eta(3)) = (1?00)$. Likewise, while computing $\widehat{\mathrm{F}}_{p,q}\,\mu(10?0)$, we consider only those cases in which $\eta(0)=\eta(1)=\eta(2)=0$ leads to the event that $(\widehat{\mathrm{F}}_{p,q}\,\eta(0), \widehat{\mathrm{F}}_{p,q}\,\eta(1), \widehat{\mathrm{F}}_{p,q}\,\eta(2), \widehat{\mathrm{F}}_{p,q}\,\eta(3)) = (10?0)$. These considerations lead to

- $$
\begin{aligned}
\widehat{\mathrm{F}}_{p,q}\,\mu(1?00) &= (1-p)rp^2\mu(000?) + (1-p)r^2p\mu(000?\{0,?\}1)\\
&\quad + (1-p)r^2(1+p-q)\mu(000?1) + C_{1?00}\\
&\geqslant (1-p)rp^2\mu(000?) + (1-p)r^2p\mu(000?\{0,?\}1) + (1-p)r^2(1+p-q)\mu(000?1); \qquad (4.52)
\end{aligned}
$$
- $$
\begin{aligned}
\widehat{\mathrm{F}}_{p,q}\,\mu(10?0) &= (1-p)p^2r\mu(000\widehat{**}) + (1-p)pr^2\mu(000\widehat{**}1) + C_{10?0} \geqslant (1-p)p^2r\mu(000\widehat{**})\\
&\quad + (1-p)pr^2\mu(000\widehat{**}1); \qquad (4.53)
\end{aligned}
$$

where $C_{1?00}$ and $C_{10?0}$ are the respective contributions from the cases where $(\eta(0),\eta(1),\eta(2)) \in \hat{\mathcal{A}}^3 \setminus \{(0,0,0)\}$. Finally, we rewrite (4.6) as

$$
\widehat{\mathrm{F}}_{p,q}\,\mu(100?) = (1-p)p^2r\mu(000\widehat{***}) + D_{100?} \geqslant (1-p)p^2r\mu(000\widehat{***}), \qquad (4.54)
$$

where $D_{100?} = qp^2r\mu(\widehat{***}\widehat{***}) + qp^2r\mu(1\{0,?\}^2\widehat{***}) + q(1-q)pr\mu(1\{0,?\}\widehat{***}) + q(1-q)^2r\mu(1\widehat{***})$.

Using (4.4), (4.54), (4.52) and (4.53), and applying the identities

1. $\mu(000\widehat{***}) = \mu(000?) + \mu(0000?) + \mu(00000?) - \mu(000?1) - \mu(000?\{0,?\}1) - \mu(0000?1)$,
2. $\mu(000\widehat{**}) = \mu(000?) - \mu(000?1) + \mu(0000?)$,
3. $\mu(000\widehat{**}1) = \mu(000?\{0,?\}1) + \mu(0000?1)$,

the sum of the last three terms, highlighted by underbraces, in (4.51) simplifies to

$$
\begin{aligned}
&- 2(q+p^2r)\,\widehat{\mathrm{F}}_{p,q}\,\mu(100?) - 2p^2r\{\widehat{\mathrm{F}}_{p,q}\,\mu(1?00) + \widehat{\mathrm{F}}_{p,q}\,\mu(10?0)\} - q\,\widehat{\mathrm{F}}_{p,q}\,\mu(?)\\
\leqslant\ &- 2(q+p^2r)(1-p)p^2r\,\underbrace{\mu(000\widehat{***})}_{\text{use (1)}} - 2p^2r\{(1-p)rp^2\mu(000?) + (1-p)r^2p\mu(000?\{0,?\}1)\\
&+ (1-p)r^2(1+p-q)\mu(000?1)\} - 2p^2r\{(1-p)p^2r\,\underbrace{\mu(000\widehat{**})}_{\text{use (2)}} + (1-p)pr^2\,\underbrace{\mu(000\widehat{**}1)}_{\text{use (3)}}\}\\
&- qr\mu(\widehat{***})\\
=\ &- 2(q+p^2r)(1-p)p^2r\\
&\underbrace{[\mu(000?) + \mu(0000?) + \mu(00000?) - \mu(000?1) - \mu(000?\{0,?\}1) - \mu(0000?1)]}_{\text{after applying (1)}}\\
&- 2p^4r^2(1-p)\mu(000?) - 2p^3r^3(1-p)\mu(000?\{0,?\}1) - 2p^2r^3(1-p)(1+p-q)\mu(000?1)\\
&- 2p^4r^2(1-p)\underbrace{[\mu(000?) - \mu(000?1) + \mu(0000?)]}_{\text{after applying (2)}}
\end{aligned}
$$

$$
\begin{aligned}
&\quad - 2p^3r^3(1-p)[\underbrace{\mu(000?\{0,?\}1) + \mu(0000?1)}_{\text{after applying (3)}}] - qr\mu(\widehat{***}) \\
&= -\big[2(q+p^2r)(1-p)p^2r + 2p^4r^2(1-p) + 2p^4r^2(1-p)\big]\mu(000?) \\
&\quad - \big[2(q+p^2r)(1-p)p^2r + 2p^4r^2(1-p)\big]\mu(0000?) - 2(q+p^2r)(1-p)p^2r\mu(00000?) \\
&\quad + \big[2(q+p^2r)(1-p)p^2r - 2p^2r^3(1-p)(1+p-q) + 2p^4r^2(1-p)\big]\mu(000?1) \\
&\quad + \big[2(q+p^2r)(1-p)p^2r - 2p^3r^3(1-p) - 2p^3r^3(1-p)\big]\mu(000?\{0,?\}1) \\
&\quad + \big[2(q+p^2r)(1-p)p^2r - 2p^3r^3(1-p)\big]\mu(0000?1) - qr\mu(\widehat{***}) \\
&\quad \text{(adding all terms involving } \mu(\mathcal{C}) \text{ for } \mathcal{C} \text{ any of } (000?), (0000?), (00000?), (000?1), \\
&\quad (000?\{0,?\}1), (0000?1)) \\
&= -[2qp^2r(1-p-q+q) + 6p^4r^2(1-p)]\mu(000?) - [2qp^2r(1-p) + 4p^4r^2(1-p)]\mu(0000?) \\
&\quad - [2qp^2r(1-p) + 2p^4r^2(1-p)]\mu(00000?) + [2qp^2r(1-p) + 4p^4r^2(1-p) \\
&\quad - 2p^2r^3(1-p)(1+p-q)]\mu(000?1) + [2qp^2r(1-p) + 2p^4r^2(1-p) \\
&\quad - 4p^3r^3(1-p)]\mu(000?\{0,?\}1) \\
&\quad + [2qp^2r(1-p) + 2p^4r^2(1-p) - 2p^3r^3(1-p)]\mu(0000?1) - qr\mu(\widehat{***}) \\
&= -[2qp^2r^2 + 2q^2p^2r + 6p^4r^2(1-p)]\mu(000?) - [2qp^2r(1-p) + 4p^4r^2(1-p)]\mu(0000?) \\
&\quad - [2qp^2r(1-p) + 2p^4r^2(1-p)]\mu(00000?) + [2qp^2r(1-p) + 4p^4r^2(1-p) \\
&\quad - 2p^2r^3(1-p)(1+p-q)]\mu(000?1) + [2qp^2r(1-p) + 2p^4r^2(1-p) \\
&\quad - 4p^3r^3(1-p)]\mu(000?\{0,?\}1) \\
&\quad + [2qp^2r(1-p) + 2p^4r^2(1-p) - 2p^3r^3(1-p)]\mu(0000?1) - qr\mu(\widehat{***}). \qquad (4.55)
\end{aligned}
$$

4.6.3 Step 3 of analyzing the effect of the adjustment in (4.49)

We begin with the following observation:

$$
\begin{aligned}
2p(p-r-p^2) &= 2p\{p-(1-p-q)-p^2\} = 2p\{2p-1+q-p^2\} \\
&= 2p\{q-(1-p)^2\} \leqslant 2p\{(1-p)-(1-p)^2\} \quad \text{(by (1.1))}; \\
&\leqslant 2p^2(1-p) \leqslant 2p(1-p) \leqslant \frac{1}{2} < 1 \text{ for all } p \in [0,1]. \qquad (4.56)
\end{aligned}
$$

This leads to the conclusion that

$$
2pqr(p-r-p^2) \leqslant qr \text{ for all } (p,q) \in \mathcal{S}, \text{ with } \mathcal{S} \text{ as defined in (1.1).} \qquad (4.57)
$$

Incorporating (4.55) into (4.51), and writing $\mu(1\widehat{***})$ as the sum of its components, i.e. $\mu(1???), \ldots, \mu(100?)$, yields (once again, in each step of the computation below, we highlight using underbraces the terms that are to be manipulated to obtain the next step):

$$
\begin{aligned}
&w_3(\widehat{\mathrm{F}}_{p,q}\,\mu) \leqslant w_3(\mu) - [p(1-r)+q]\mu(10?) - r(4q^2+4qp^2-2q^3+2q^2p)[\mu(1???) \\
&+ \mu(1?0?) + \mu(10??) + \mu(100?)] - r(4q^2+4qp^2-2q^3+2q^2p)[\mu(1?00) + \mu(10?0) \\
&+ \mu(1??0)] - 2r(1-p^2)\mu(1??1) - r(1-p)\mu(1?1) - [p(1-r)+q][\mu(0?) + \mu(00?)] \\
&- pr^2(1+2q)\mu(1\widehat{***}1) \underbrace{+2pqr(p-r-p^2)\mu(\widehat{***})} - pr^2\{1+2q(1-p)\}\mu(?000?) \\
&- 6pqr^2(1-p)\mu(0000?) + [2pqr^2\{4p-2+p^2\} + 6p^4r^2 - 4p^4r^2(p+q)]\mu(000?) \\
&+ 2p^2r^2(1-p)(1-pr)\mu(00000?) - 2p^3r^3(1-p)\{\mu(10000?) + \mu(?0000?)\} \\
&- 2pr^2\{1+q-p+2p^2-2p^2r(1-p)+pq(1-p)-p^3\}\mu(000?1) \\
&- 4(1-p)r^2q(r-p)\mu(000?1) - 2p^2r^2(1-p)(1-2pr)\mu(000?\{0,?\}1) \\
&- 2p^2r^2(1-p)(1-pr)\mu(0000?1) - 2p^2qr^2\mu(1000?)
\end{aligned}
$$

$$
\begin{aligned}
&- 4qp^2r^2[\mu(1\{0,?\}\widehat{***}) - \mu(1000?)] - 2pqr^2(2-2q+p)\mu(1\widehat{***})\\
&- [2qp^2r^2 + 2q^2p^2r + 6p^4r^2(1-p)]\mu(000?) - [2qp^2r(1-p) + 4p^4r^2(1-p)]\mu(0000?)\\
&- [2qp^2r(1-p) + 2p^4r^2(1-p)]\mu(00000?) + [2qp^2r(1-p) + 4p^4r^2(1-p)\\
&- 2p^2r^3(1-p)(1+p-q)]\mu(000?1) + [2qp^2r(1-p) + 2p^4r^2(1-p)\\
&- 4p^3r^3(1-p)]\mu(000?\{0,?\}1)\\
&+ [2qp^2r(1-p) + 2p^4r^2(1-p) - 2p^3r^3(1-p)]\mu(0000?1) \underbrace{-qr\mu(\widehat{***})}\\
= {}& w_3(\mu) - [p(1-r)+q]\mu(10?) - r(4q^2 + 4qp^2 - 2q^3 + 2q^2p)[\mu(1???) + \mu(1?0?)\\
&+ \mu(10??) + \mu(100?)] - r(4q^2 + 4qp^2 - 2q^3 + 2q^2p)[\mu(1?00) + \mu(10?0) + \mu(1??0)]\\
&- 2r(1-p^2)\mu(1??1) - r(1-p)\mu(1?1) - [p(1-r)+q][\mu(0?) + \mu(00?)]\\
&- pr^2(1+2q)\mu(1\widehat{***}1) \underbrace{-qr[1-2p(p-r-p^2)]\mu(\widehat{***})}_{\text{non-positive by (4.57)}} - pr^2\{1+2q(1-p)\}\mu(?000?)\\
&- 6pqr^2(1-p)\mu(0000?) + [2pqr^2\{4p-2+p^2\} + 6p^4r^2 - 4p^4r^2(p+q)]\mu(000?)\\
&+ 2p^2r^2(1-p)(1-pr)\mu(00000?) - 2p^3r^3(1-p)\{\mu(10000?) + \mu(?0000?)\}\\
&- 2pr^2\{1+q-p+2p^2-2p^2r(1-p)+pq(1-p)-p^3\}\mu(000?1)\\
&- 4(1-p)r^2q(r-p)\mu(000?1) - 2p^2r^2(1-p)(1-2pr)\mu(000?\{0,?\}1)\\
&- 2p^2r^2(1-p)(1-pr)\mu(0000?1) - 2p^2qr^2\mu(1000?)\\
&- 4qp^2r^2[\mu(1\{0,?\}\widehat{***}) - \mu(1000?)] - 2pqr^2(2-2q+p)\mu(1\widehat{***})\\
&- [2qp^2r^2 + 2q^2p^2r + 6p^4r^2(1-p)]\mu(000?) - [2qp^2r(1-p) + 4p^4r^2(1-p)]\mu(0000?)\\
&- [2qp^2r(1-p) + 2p^4r^2(1-p)]\mu(00000?) + [2qp^2r(1-p) + 4p^4r^2(1-p)\\
&- 2p^2r^3(1-p)(1+p-q)]\mu(000?1)\\
&+ [2qp^2r(1-p) + 2p^4r^2(1-p) - 4p^3r^3(1-p)]\mu(000?\{0,?\}1)\\
&+ [2qp^2r(1-p) + 2p^4r^2(1-p) - 2p^3r^3(1-p)]\mu(0000?1)\\
= {}& w_3(\mu) - [p(1-r)+q]\mu(10?) - r(4q^2 + 4qp^2 - 2q^3 + 2q^2p)[\mu(1???) + \mu(1?0?)\\
&+ \mu(10??) + \mu(100?)] \underbrace{-r(4q^2 + 4qp^2 - 2q^3 + 2q^2p)[\mu(1?00) + \mu(10?0) + \mu(1??0)]}\\
&- 2r(1-p^2)\mu(1??1) - r(1-p)\mu(1?1) \underbrace{-[p(1-r)+q][\mu(0?) + \mu(00?)]}\\
&- pr^2(1+2q)\mu(1\widehat{***}1) - qr[1-2p(p-r-p^2)]\mu(\widehat{***}) - pr^2\{1+2q(1-p)\}\mu(?000?)\\
&- 6pqr^2(1-p)\mu(0000?) \underbrace{+[2pqr^2\{4p-2+p^2\} + 6p^4r^2 - 4p^4r^2(p+q)]\mu(000?)}\\
&\underbrace{+2p^2r^2(1-p)(1-pr)\mu(00000?)} - 2p^3r^3(1-p)\{\mu(10000?) + \mu(?0000?)\}\\
&\underbrace{-2pr^2\{1+q-p+2p^2-2p^2r(1-p)+pq(1-p)-p^3\}\mu(000?1) - 4(1-p)r^2q(r-p)\mu(000?1)}\\
&\underbrace{-2p^2r^2(1-p)(1-2pr)\mu(000?\{0,?\}1) - 2p^2r^2(1-p)(1-pr)\mu(0000?1)} - 2p^2qr^2\mu(1000?)\\
&- 4qp^2r^2[\mu(1\{0,?\}\widehat{***}) - \mu(1000?)] - 2pqr^2(2-2q+p)\mu(1\widehat{***})\\
&\underbrace{-[2qp^2r^2 + 2q^2p^2r + 6p^4r^2(1-p)]\mu(000?) - [2qp^2r(1-p) + 4p^4r^2(1-p)]\mu(0000?)}\\
&\underbrace{-[2qp^2r(1-p) + 2p^4r^2(1-p)]\mu(00000?) + [2qp^2r(1-p) + 4p^4r^2(1-p)}\\
&\underbrace{-2p^2r^3(1-p)(1+p-q)]\mu(000?1) + [2qp^2r(1-p) + 2p^4r^2(1-p) - 4p^3r^3(1-p)]\mu(000?\{0,?\}1)}\\
&\underbrace{+[2qp^2r(1-p) + 2p^4r^2(1-p) - 2p^3r^3(1-p)]\mu(0000?1)}. \qquad (4.58)
\end{aligned}
$$

We collect all the terms that have been highlighted in the *last step* of (4.58) using underbraces, and denote their sum by C_2, whereas the sum of the remaning terms is denoted C_1. In other words,

$$\begin{aligned} C_1 = w_3(\mu) &- [p(1-r)+q]\mu(10?) - r(4q^2+4qp^2-2q^3+2q^2p)[\mu(1???)+\mu(1?0?) \\ &+ \mu(10??)+\mu(100?)] - 2r(1-p^2)\mu(1??1) - r(1-p)\mu(1?1) - pr^2(1+2q)\mu(1\widehat{***}1) \\ &- qr[1-2p(p-r-p^2)]\mu(\widehat{***}) - pr^2\{1+2q(1-p)\}\mu(?000?) - 6pqr^2(1-p)\mu(0000?) \\ &- 2p^3r^3(1-p)\{\mu(10000?)+\mu(?0000?)\} - 2p^2qr^2\mu(1000?) \\ &- 4qp^2r^2[\mu(1\{0,?\}\widehat{***}) - \mu(1000?)] - 2pqr^2(2-2q+p)\mu(1\widehat{***}), \end{aligned} \tag{4.59}$$

and

$$\begin{aligned} C_2 = &-r(4q^2+4qp^2-2q^3+2q^2p)[\mu(1?00)+\mu(10?0)+\mu(1??0)] \\ &- [p(1-r)+q][\mu(0?)+\mu(00?)] + [2pqr^2\{4p-2+p^2\} \\ &+ 6p^4r^2 - 4p^4r^2(p+q)]\mu(000?) + 2p^2r^2(1-p)(1-pr)\mu(00000?) \\ &- 2pr^2\{1+q-p+2p^2-2p^2r(1-p)+pq(1-p)-p^3\}\mu(000?1) \\ &- 4(1-p)r^2q(r-p)\mu(000?1) \\ &- 2p^2r^2(1-p)(1-2pr)\mu(000?\{0,?\}1) - 2p^2r^2(1-p)(1-pr)\mu(0000?1) \\ &- [2qp^2r^2+2q^2p^2r+6p^4r^2(1-p)]\mu(000?) - [2qp^2r(1-p)+4p^4r^2(1-p)]\mu(0000?) \\ &- [2qp^2r(1-p)+2p^4r^2(1-p)]\mu(00000?) + [2qp^2r(1-p)+4p^4r^2(1-p) \\ &- 2p^2r^3(1-p)(1+p-q)]\mu(000?1) + [2qp^2r(1-p)+2p^4r^2(1-p)-4p^3r^3(1-p)] \\ &\mu(000?\{0,?\}1) + [2qp^2r(1-p)+2p^4r^2(1-p)-2p^3r^3(1-p)]\mu(0000?1). \end{aligned} \tag{4.60}$$

It is worthwhile to note that all terms in C_1, apart from $w_3(\mu)$, are non-positive.

Our task, now, is to see if we can turn each of the non-negative terms in C_2 non-positive by making use of the existing non-positive terms in (4.60). In the long computation that follows, the terms that are being dealt with in each step will be highlighted using underbraces. First, we note that, in

$$\begin{aligned} C_2 = &-r(4q^2+4qp^2-2q^3+2q^2p)[\mu(1?00)+\mu(10?0)+\mu(1??0)] \\ &- [p(1-r)+q][\mu(0?)+\mu(00?)] \\ &+ [2pqr^2\{4p-2+p^2\}+6p^4r^2-4p^4r^2(p+q)]\mu(000?) \\ &+ 2p^2r^2(1-p)(1-pr)\mu(00000?) \\ &- 2pr^2\{1+q-p+2p^2-2p^2r(1-p)+pq(1-p)-p^3\}\mu(000?1) \\ &- 4(1-p)r^2q(r-p)\mu(000?1) - 2p^2r^2(1-p)(1-2pr)\mu(000?\{0,?\}1) \\ &- 2p^2r^2(1-p)(1-pr)\mu(0000?1) \\ &- [2qp^2r^2+2q^2p^2r+6p^4r^2(1-p)]\mu(000?) - [2qp^2r(1-p)+4p^4r^2(1-p)]\mu(0000?) \\ &- [2qp^2r(1-p)+2p^4r^2(1-p)]\mu(00000?) + [2qp^2r(1-p)+4p^4r^2(1-p) \\ &- 2p^2r^3(1-p)(1+p-q)]\mu(000?1) + [2qp^2r(1-p)+2p^4r^2(1-p) \\ &- 4p^3r^3(1-p)]\mu(000?\{0,?\}1) + [2qp^2r(1-p)+2p^4r^2(1-p)-2p^3r^3(1-p)]\mu(0000?1) \\ = &\underbrace{-r(4q^2+4qp^2-2q^3+2q^2p)\mu(1?00) - 2pr^2\{1+q-p+2p^2-2p^2r(1-p)+pq(1-p)-p^3\}\mu(000?1)}_{A_1} \\ &\underbrace{-4(1-p)r^2q(r-p)\mu(000?1) + [2qp^2r(1-p)+4p^4r^2(1-p)-2p^2r^3(1-p)(1+p-q)]\mu(000?1)}_{A_1} \\ &\underbrace{-r(4q^2+4qp^2-2q^3+2q^2p)[\mu(10?0)+\mu(1??0)] - 2p^2r^2(1-p)(1-2pr)\mu(000?\{0,?\}1)}_{A_2} \\ &\underbrace{+[2qp^2r(1-p)+2p^4r^2(1-p)-4p^3r^3(1-p)]\mu(000?\{0,?\}1)}_{A_2} \underbrace{-[2qp^2r(1-p)+4p^4r^2(1-p)]\mu(0000?)}_{A_3} \\ &\underbrace{+[2qp^2r(1-p)+2p^4r^2(1-p)-2p^3r^3(1-p)]\mu(0000?1) - 2p^2r^2(1-p)(1-pr)\mu(0000?1)}_{A_3} \end{aligned}$$

$$
\begin{aligned}
&- [p(1-r)+q][\mu(0?)+\mu(00?)] + [2pqr^2\{4p-2+p^2\} + 6p^4r^2 - 4p^4r^2(p+q)]\mu(000?)\\
&+ 2p^2r^2(1-p)(1-pr)\mu(00000?) - [2qp^2r^2 + 2q^2p^2r + 6p^4r^2(1-p)]\mu(000?)\\
&- [2qp^2r(1-p) + 2p^4r^2(1-p)]\mu(00000?). \qquad (4.61)
\end{aligned}
$$

In what follows, we let A_i, for $i \in \{1,2,3\}$, denote the sum of the terms that have been highlighted in (4.61) using underbraces tagged A_i, and we simplify each A_i separately before incorporating these simplified expressions back into (4.61).

First, we simplify A_1, using the simple inequality $\mu(1?00) \leqslant \mu(1?000)$, as follows:

$$
\begin{aligned}
A_1 =\ & \underbrace{-r(4q^2+4qp^2-2q^3+2q^2p)\mu(1?00)}_{\text{split into two parts as shown below}}\\
& - 2pr^2\{1+q-p+2p^2-2p^2r(1-p)+pq(1-p)-p^3\}\mu(000?1)\\
& + \{2qp^2r(1-p)+4p^4r^2(1-p)-2p^2r^3(1-p)(1+p-q)\}\mu(000?1)\\
& - 4(1-p)r^2q(r-p)\mu(000?1)\\
=\ & \underbrace{-r(4q^2+2qp^2(1+p)-2q^3+2q^2p)\mu(1?00) - 2qp^2r(1-p)\mu(1?00)}_{\text{after splitting into two parts}}\\
& - 2pr^2\{1+q-p+2p^2-2p^2(1-p-q)(1-p)+pq(1-p)-p^3\}\mu(000?1)\\
& + \{2qp^2r(1-p)+4p^4r^2(1-p)-2p^2r^3(1-p)(1+p-q)\}\mu(000?1)\\
& - 4(1-p)r^2q(r-p)\mu(000?1)\\
\leqslant\ & - r(4q^2+2qp^2(1+p)-2q^3+2q^2p)\mu(1?00) \underbrace{-2qp^2r(1-p)\mu(1?000)}_{\text{cancel out}}\\
& - 2pr^2\{1+q-p+2p^2-2p^2(1-p)^2+2p^2q(1-p)+pq(1-p)-p^3\}\mu(000?1)\\
& \underbrace{+2qp^2r(1-p)\mu(000?1)}_{\text{cancel out}}\\
& + \{4p^4r^2(1-p)-2p^2r^3(1-p)(1+p-q)\}\mu(000?1) - 4(1-p)r^2q(r-p)\mu(000?1)\\
=\ & - r(4q^2+2qp^2(1+p)-2q^3+2q^2p)\mu(1?00) - 2pr^2\{1+q-p+2p^2-2p^2+4p^3-2p^4\\
& + 2p^2q(1-p)+pq(1-p)-p^3\}\mu(000?1) + \{4p^4r^2(1-p)\\
& - 2p^2r^3(1-p)(1+p-q)\}\mu(000?1) - 4(1-p)r^2q(r-p)\mu(000?1)\\
=\ & - r(4q^2+2qp^2(1+p)-2q^3+2q^2p)\mu(1?00)\\
& - 2pr^2\{1+q-p+p^3+2p^2q(1-p)+pq(1-p)\}\mu(000?1)\\
& - 2pr^2(2p^3-2p^4)\mu(000?1) + \{4p^4r^2(1-p)-2p^2r^3(1-p)(1+p-q)\}\mu(000?1)\\
& - 4(1-p)r^2q(r-p)\mu(000?1)\\
=\ & - r(4q^2+2qp^2(1+p)-2q^3+2q^2p)\mu(1?00)\\
& - 2pr^2\{1+q-p+p^3+2p^2q(1-p)+pq(1-p)\}\mu(000?1)\\
& - 4p^4r^2(1-p)\mu(000?1) + 4p^4r^2(1-p)\mu(000?1) - 2p^2r^3(1-p)(1+p-q)\mu(000?1)\\
& - 4(1-p)r^2q(r-p)\mu(000?1)\\
=\ & - r(4q^2+2qp^2(1+p)-2q^3+2q^2p)\mu(1?00)\\
& - 2pr^2\{1+q-p+p^3+2p^2q(1-p)+pq(1-p)\}\mu(000?1)\\
& - 2p^2r^3(1-p)(1+p-q)\mu(000?1) - 4(1-p)r^2q(r-p)\mu(000?1). \qquad (4.62)
\end{aligned}
$$

Lemma 4.2. *The coefficient of* $\mu(000?1)$ *in* (4.62) *is bounded above by* $-2p^2r^3(1-p)(1+p-q)$ *for all values of* $(p,q) \in \mathcal{S}$*, where* $\mathcal{S}$ *is as defined in* (1.1).

Proof. As long as $r-p \geqslant 0$, i.e. $2p+q \leqslant 1$, the coefficient of $\mu(000?1)$ in (4.62) is the sum of three non-positive quantities, namely $-2pr^2\{1+q-p+p^3+2p^2q(1-p)+pq(1-p)\}$, $-2p^2r^3(1-p)(1+p-q)$ and $-4(1-p)r^2q(r-p)$, and hence it is obviously bounded above by the second of these non-positive quantities. When $r - p < 0$, the quantity $-4(1-p)r^2q(r-p)$ is actually non-negative, and we negate it by observing that

$$\begin{aligned}
&-2pr^2\{1+q-p+p^3+2p^2q(1-p)+pq(1-p)\}+4(1-p)r^2q(p-r)\\
&=-2pr^2\{1+q-p+p^3+2p^2q(1-p)+pq(1-p)\}+4(1-p)r^2qp-4(1-p)r^3q\\
&\leqslant -2pr^2\{1+q-p+p^3+2p^2q(1-p)+pq(1-p)\}+4(1-p)r^2qp\\
&=-2pr^2\{1+q-p+p^3+2p^2q(1-p)+pq(1-p)-2q(1-p)\}\\
&=-2pr^2\{1+q-p+p^3+2p^2q(1-p)+pq(1-p)-2q+2pq\}\\
&=-2pr^2\{r+p^3+2p^2q(1-p)+pq(1-p)+2pq\},
\end{aligned}$$

which is non-positive. This again leads to the conclusion that the combined coefficient of $\mu(000?1)$ in (4.62) is bounded above by $-2p^2r^3(1-p)(1+p-q)$. ☐

Next, using the simple inequalities $\mu(000??1) \leqslant \mu(1??0)$ and $\mu(000?01) \leqslant \mu(10?0)$, we simplify A_2 (recall A_2 from (4.61)) as follows:

$$\begin{aligned}
A_2 = &-r(4q^2+4qp^2-2q^3+2q^2p)\{\mu(1??0)+\mu(10?0)\}\\
&-2p^2r^2(1-p)(1-2pr)\mu(000?\{0,?\}1)\\
&+\{2qp^2r(1-p)+2p^4r^2(1-p)-4p^3r^3(1-p)\}\mu(000?\{0,?\}1)\\
\leqslant\ &\underbrace{-r\{4q^2+4qp^2-2q^3+2q^2p\}\{\mu(1??0)+\mu(10?0)\}}_{(1)}\\
&\underbrace{-2p^2r^2(1-p)(1-2pr)\mu(000?\{0,?\}1)}_{(2)}\\
&\underbrace{+2qp^2r(1-p)\{\mu(1??0)+\mu(10?0)\}}_{\text{combine with (1)}}\\
&\underbrace{+2p^4r^2(1-p)\mu(000?\{0,?\}1)-4p^3r^3(1-p)\mu(000?\{0,?\}1)}_{\text{combine with (2)}}\\
=\ &\underbrace{-r\{4q^2+4qp^2-2q^3+2q^2p-2qp^2(1-p)\}\{\mu(1??0)+\mu(10?0)\}}_{\text{(1) after combination}}\\
&\underbrace{-2p^2r^2(1-p)[1-2pr-p^2+2pr]\mu(000?\{0,?\}1)}_{\text{(2) after combination}}\\
=\ &-r\{4q^2+2qp^2(1+p)-2q^3+2q^2p\}\{\mu(1??0)+\mu(10?0)\}\\
&-2p^2r^2(1-p)(1-p^2)\mu(000?\{0,?\}1).
\end{aligned} \tag{4.63}$$

Finally, using the simple inequality $\mu(0000?1) \leqslant \mu(0000?)$, we simplify A_3 (recall A_3 from (4.61)) as follows:

$$\begin{aligned}
A_3 = &-\{2qp^2r(1-p)+4p^4r^2(1-p)\}\mu(0000?)-2p^2r^2(1-p)(1-pr)\mu(0000?1)\\
&+\{2qp^2r(1-p)+2p^4r^2(1-p)-2p^3r^3(1-p)\}\mu(0000?1)\\
=\ &-2qp^2r(1-p)\mu(0000?)+2qp^2r(1-p)\mu(0000?1)-4p^4r^2(1-p)\mu(0000?)\\
&-[2p^2r^2(1-p)(1-pr)-2p^4r^2(1-p)+2p^3r^3(1-p)]\mu(0000?1)\\
\leqslant\ &-2p^2r^2(1-p)[1-pr-p^2+pr]\mu(0000?1)-4p^4r^2(1-p)\mu(0000?)\\
=\ &-2p^2r^2(1-p)(1-p^2)\mu(0000?1)-4p^4r^2(1-p)\mu(0000?).
\end{aligned} \tag{4.64}$$

Incorporating the expressions obtained from (4.62) (in conjunction with the bound obtained from Lemma 4.2), (4.63) and (4.64) into (4.61) yields (again, in each step, we highlight using underbraces the terms to be combined to obtain the next step):

$$C_2 \leqslant \underbrace{-r(4q^2+2qp^2(1+p)-2q^3+2q^2p)\mu(1?00)-2p^2r^3(1-p)(1+p-q)\mu(000?1)}_{\text{substituting from Lemma 4.2 and (4.62)}}$$

$$
\begin{aligned}
&\underbrace{-r(4q^2+2qp^2(1+p)-2q^3+2q^2p)[\mu(10?0)+\mu(1??0)]-2p^2r^2(1-p)(1-p^2)\mu(000?\{0,?\}1)}_{\text{substituting from (4.63)}}\\
&\underbrace{-2p^2r^2(1-p)(1-p^2)\mu(0000?1)-4p^4r^2(1-p)\mu(0000?)}_{\text{substituting from (4.64)}}-[p(1-r)+q][\mu(0?)+\mu(00?)]\\
&+[2pqr^2\{4p-2+p^2\}+6p^4r^2-4p^4r^2(p+q)]\mu(000?)+2p^2r^2(1-p)(1-pr)\mu(00000?)\\
&-[2qp^2r^2+2q^2p^2r+6p^4r^2(1-p)]\mu(000?)-[2qp^2r(1-p)+2p^4r^2(1-p)]\mu(00000?)\\
=\;&-r(4q^2+2qp^2(1+p)-2q^3+2q^2p)\{\mu(1?00)+\mu(10?0)+\mu(1??0)\}\\
&-[p(1-r)+q][\mu(0?)+\mu(00?)]\underbrace{+[2pqr^2\{4p-2+p^2\}+6p^4r^2-4p^4r^2(p+q)]\mu(000?)}\\
&+2p^2r^2(1-p)(1-pr)\mu(00000?)-2p^2r^3(1-p)(1+p-q)\mu(000?1)\\
&-2p^2r^2(1-p)(1-p^2)\mu(000?\{0,?\}1)-2p^2r^2(1-p)(1-p^2)\mu(0000?1)\\
&\underbrace{-[2qp^2r^2+2q^2p^2r+6p^4r^2(1-p)]\mu(000?)}-4p^4r^2(1-p)\mu(0000?)\\
&-[2qp^2r(1-p)+2p^4r^2(1-p)]\mu(00000?)\\
=\;&-r(4q^2+2qp^2(1+p)-2q^3+2q^2p)\{\mu(1?00)+\mu(10?0)+\mu(1??0)\}\\
&\underbrace{-[p(1-r)+q][\mu(0?)+\mu(00?)]}+[2pqr^2\{3p-2+p^2\}-2q^2p^2r\underbrace{+2p^5r^2}-4p^4qr^2]\mu(000?)\\
&\underbrace{+2p^2r^2(1-p)(1-pr)\mu(00000?)}-2p^2r^3(1-p)(1+p-q)\mu(000?1)\\
&-2p^2r^2(1-p)(1-p^2)\mu(000?\{0,?\}1)-2p^2r^2(1-p)(1-p^2)\mu(0000?1)\\
&\underbrace{-4p^4r^2(1-p)\mu(0000?)-[2qp^2r(1-p)+2p^4r^2(1-p)]\mu(00000?)}. \qquad (4.65)
\end{aligned}
$$

We combine the terms highlighted by underbrace in the last step of (4.65) as follows:

$$
\begin{aligned}
&2p^2r^2(1-p)(1-pr)\mu(00000?)-[2qp^2r(1-p)+2p^4r^2(1-p)]\mu(00000?)+2p^5r^2\mu(000?)\\
&-4p^4r^2(1-p)\mu(0000?)-[p(1-r)+q][\mu(0?)+\mu(00?)]\\
=\;&2p^2r^2(1-p)(1-pr)\mu(00000?)-2p^4r^2(1-p)\mu(00000?)-4p^4r^2(1-p)\mu(0000?)\\
&+2p^5r^2\mu(000?)-2qp^2r(1-p)\mu(00000?)-[p(1-r)+q][\mu(0?)+\mu(00?)]\\
&\text{(rearranging the terms of the previous expression)}\\
\leqslant\;&2p^2r^2(1-p)(1-pr)\mu(00000?)-2p^4r^2(1-p)\mu(00000?)-4p^4r^2(1-p)\mu(00000?)\\
&+2p^5r^2\mu(000?)-2qp^2r(1-p)\mu(00000?)-[p^2+q(1+p)][\mu(0?)+\mu(00?)]\\
&\text{(using the inequality } \mu(0000?)\geqslant\mu(00000?)\text{)}\\
=\;&2p^2r^2(1-p)(1-pr-3p^2)\mu(00000?)+2p^5r^2\mu(000?)-p^2\{\mu(0?)+\mu(00?)\}\\
&-2qp^2r(1-p)\mu(00000?)-q(1+p)\{\mu(0?)+\mu(00?)\}. \qquad (4.66)
\end{aligned}
$$

Lemma 4.3. *The sum in* (4.66) *is non-positive for all* $(p,q)\in\mathcal{S}$*, where* $\mathcal{S}$ *is as defined in* (1.1)*. Moreover, it is bounded above by* $-2qp^2r(1-p)\mu(00000?)-q(1+p)\{\mu(0?)+\mu(00?)\}$.

Proof. In this proof, we make use of the inequalities $\mu(0?)\geqslant\mu(00?)\geqslant\mu(000?)\geqslant\mu(00000?)$. If $1-pr-3p^2\leqslant 0$, each of the terms in (4.66), except $2p^5r^2\mu(000?)$, is non-positive, and we tackle this single non-negative term as follows:

$$2p^5r^2\mu(000?)-p^2\{\mu(0?)+\mu(00?)\}=2p^5r^2\mu(000?)-2p^2\mu(000?)=-2p^2(1-p^3r^2)\mu(000?), \qquad (4.67)$$

and every term in the above sum is non-positive. If $1-pr-3p^2>0$, both the first and the second terms of (4.66) are non-negative, while the rest are non-positive, and these non-negative terms are taken care of as follows:

$$2p^2r^2(1-p)(1-pr-3p^2)\mu(00000?)+2p^5r^2\mu(000?)-p^2\{\mu(0?)+\mu(00?)\}$$

$$
\begin{aligned}
&\leqslant 2p^2r^2(1-p)(1-pr-3p^2)\mu(000?)+2p^5r^2\mu(000?)-2p^2\mu(000?)\\
&=2p^2r^2[(1-p)(1-pr-3p^2)+p^3]\mu(000?)-2p^2\mu(000?)\\
&=2p^2r^2[1-p(1-p^2)-pr(1-p)-3p^2(1-p)]\mu(000?)-2p^2\mu(000?)\\
&=2p^2[r^2\{1-p(1-p^2)-pr(1-p)-3p^2(1-p)\}-1]\mu(000?),
\end{aligned}\tag{4.68}
$$

and once again, we see that each term in the above expression is non-positive, since each of r^2 and $1-p(1-p^2)-pr(1-p)-3p^2(1-p)$ is bounded above by 1. From the final upper bounds in (4.67) and (4.68), we deduce that the expression in (4.66) is bounded above by $-2qp^2r(1-p)\mu(00000?)-q(1+p)\{\mu(0?)+\mu(00?)\}$. □

Incorporating (4.66) into (4.65) and using Lemma 4.3, we see that C_2 can be bounded above as follows:

$$
\begin{aligned}
C_2 \leqslant\; & -r(4q^2+2qp^2(1+p)-2q^3+2q^2p)\{\mu(1?00)+\mu(10?0)+\mu(1??0)\}\\
& \underbrace{-q(1+p)[\mu(0?)+\mu(00?)]}_{\text{split this into 2 parts, as below}}+[\underbrace{2pqr^2\{3p-2+p^2\}}_{\text{consider this separately}}-2q^2p^2r-4p^4qr^2]\mu(000?)\\
& -2qp^2r(1-p)\mu(00000?)-2p^2r^3(1-p)(1+p-q)\mu(000?1)\\
& \underbrace{-2p^2r^2(1-p)(1-p^2)\mu(000?\{0,?\}1)-2p^2r^2(1-p)(1-p^2)\mu(0000?1)}_{\text{combine these two}}\\
=\; & -r(4q^2+2qp^2(1+p)-2q^3+2q^2p)\{\mu(1?00)+\mu(10?0)+\mu(1??0)\}\\
& -q(1+p-pr)[\mu(0?)+\mu(00?)]\quad \underbrace{-qpr[\mu(0?)+\mu(00?)]}_{\text{bounded above by }-2pqr\mu(000?)}\quad +2pqr^2\{3p-2+p^2\}\mu(000?)\\
& -[2q^2p^2r+4p^4qr^2]\mu(000?)-2qp^2r(1-p)\mu(00000?)\\
& -2p^2r^3(1-p)(1+p-q)\mu(000?1)-2p^2r^2(1-p)(1-p^2)\mu(000\widehat{**}1)\\
\leqslant\; & -r(4q^2+2qp^2(1+p)-2q^3+2q^2p)\{\mu(1?00)+\mu(10?0)+\mu(1??0)\}\\
& -q(1+p-pr)[\mu(0?)+\mu(00?)]\underbrace{-2qpr\mu(000?)+2pqr^2\{3p-2+p^2\}\mu(000?)}\\
& -[2q^2p^2r+4p^4qr^2]\mu(000?)-2qp^2r(1-p)\mu(00000?)\\
& -2p^2r^3(1-p)(1+p-q)\mu(000?1)-2p^2r^2(1-p)(1-p^2)\mu(000\widehat{**}1)\\
=\; & -r(4q^2+2qp^2(1+p)-2q^3+2q^2p)\{\mu(1?00)+\mu(10?0)+\mu(1??0)\}\\
& -q(1+p-pr)[\mu(0?)+\mu(00?)]+\underbrace{[-2qpr+4p^2qr^2+2pqr^2\{p-2+p^2\}]\mu(000?)}\\
& -[2q^2p^2r+4p^4qr^2]\mu(000?)-2qp^2r(1-p)\mu(00000?)\\
& -2p^2r^3(1-p)(1+p-q)\mu(000?1)-2p^2r^2(1-p)(1-p^2)\mu(000\widehat{**}1)\\
=\; & -r(4q^2+2qp^2(1+p)-2q^3+2q^2p)\{\mu(1?00)+\mu(10?0)+\mu(1??0)\}\\
& -q(1+p-pr)[\mu(0?)+\mu(00?)]\underbrace{-2qpr[1-2pr]\mu(000?)-2pqr^2\{2-p-p^2\}\mu(000?)}\\
& -[2q^2p^2r+4p^4qr^2]\mu(000?)-2qp^2r(1-p)\mu(00000?)\\
& -2p^2r^3(1-p)(1+p-q)\mu(000?1)-2p^2r^2(1-p)(1-p^2)\mu(000\widehat{**}1)\\
=\; & -r(4q^2+2qp^2(1+p)-2q^3+2q^2p)\{\mu(1?00)+\mu(10?0)+\mu(1??0)\}\\
& -q(1+p-pr)[\mu(0?)+\mu(00?)]\underbrace{-2qpr[1-2pr]\mu(000?)}-2pqr^2(2+p)(1-p)\mu(000?)\\
& -[2q^2p^2r+4p^4qr^2]\mu(000?)-2qp^2r(1-p)\mu(00000?)\\
& -2p^2r^3(1-p)(1+p-q)\mu(000?1)-2p^2r^2(1-p)(1-p^2)\mu(000\widehat{**}1).
\end{aligned}\tag{4.69}
$$

We note that, by the AM-GM inequality, $2pr\leqslant p^2+r^2\leqslant p+r\leqslant 1$, so that the underbraced term, and consequently, each term in the final expression of (4.69), is non-positive. Note that (4.69) also provides the final, simplified expression for C_2. From (4.59) and (4.69), we see that (4.58) now transforms into the weight function inequality

$$
\begin{aligned}
& w_3(\widehat{\mathrm{F}}_{p,q}\,\mu) \leqslant w_3(\mu) - [p(1-r)+q]\mu(10?) - r(4q^2+4qp^2-2q^3+2q^2p)[\mu(1???)+\mu(1?0?) \\
& + \mu(10??)+\mu(100?)] - 2r(1-p^2)\mu(1??1) - r(1-p)\mu(1?1) - pr^2(1+2q)\mu(1\widehat{***}1) \\
& - qr[1-2p(p-r-p^2)]\mu(\widehat{***}) - pr^2\{1+2q(1-p)\}\mu(?000?) - 6pqr^2(1-p)\mu(0000?) \\
& - 2p^3r^3(1-p)\{\mu(10000?)+\mu(?0000?)\} - 2p^2qr^2\mu(1000?) \\
& - 4qp^2r^2[\mu(1\{0,?\}\widehat{***}) - \mu(1000?)] - 2pqr^2(2-2q+p)\mu(1\widehat{***}) \\
& - r(4q^2+2qp^2(1+p)-2q^3+2q^2p)\{\mu(1?00)+\mu(10?0) \\
& + \mu(1??0)\} - q(1+p-pr)[\mu(0?)+\mu(00?)] - 2qpr[1-2pr]\mu(000?) \\
& - 2pqr^2(2+p)(1-p)\mu(000?) - [2q^2p^2r+4p^4qr^2]\mu(000?) - 2qp^2r(1-p)\mu(00000?) \\
& - 2p^2r^3(1-p)(1+p-q)\mu(000?1) - 2p^2r^2(1-p)(1-p^2)\mu(000\widehat{**}1).
\end{aligned} \tag{4.70}
$$

Finally, we have achieved a weight function inequality that is of the form presented in (4.2).

4.7 The desired conclusion

From (4.10), (4.17), (4.23) and (4.49), the final weight function turns out to be

$$
\begin{aligned}
w_3(\mu) &= \mu(?) + 2\mu(0?) - \mu(?0?) + 2\mu(100?) - p(p+q)\mu(?) - [2pr\{\mu(1?)+\mu(10?)\} \\
&\quad + 2p^2r\{\mu(1??)+\mu(1?0?)+\mu(10??)\} + 4r\mu(1?01) + 2p\mu(100?)] - 2(q+p^2r)\mu(100?) \\
&\quad - 2p^2r\{\mu(1?00)+\mu(10?0)\} - q\mu(?) \\
&= (1-p^2-pq-q)\mu(?) + 2\mu(0?) - \mu(?0?) + [2-2p-2q-2p^2r]\mu(100?) - 2pr\mu(1?) \\
&\quad - 2pr\mu(10?) - 2p^2r\{\mu(1??)+\mu(1?0?)+\mu(10??)\} - 4r\mu(1?01) \\
&\quad - 2p^2r\{\mu(1?00)+\mu(10?0)\},
\end{aligned}
$$

which matches with the weight function stated in (4.3).

We now wish to draw the conclusion, using (4.70), that $\mu(?) = 0$ when μ is a translation-invariant and reflection-invariant stationary distribution for $\widehat{\mathrm{F}}_{p,q}$ (recall Lemma 4.1). Note that this conclusion, i.e. that $\mu(?) = 0$, is immediate when $r = 0$, i.e. $p + q = 1$, so that henceforth, we only consider $r > 0$, i.e. $p + q < 1$.

To begin with, we note that when μ is stationary for $\widehat{\mathrm{F}}_{p,q}$, we have $\widehat{\mathrm{F}}_{p,q}\,\mu = \mu$, so that $w_3(\widehat{\mathrm{F}}_{p,q}\,\mu) = w_3(\mu)$. When $q > 0$, we conclude that the coefficient $-qr[1-2p(p-r-p^2)]$ of $\mu(\widehat{***})$ in (4.70) is non-zero (see (4.56) that shows why $1-2p(p-r-p^2)$ is strictly positive). The last two sentences and (4.70) together imply that $\mu(\widehat{***}) = 0$ when μ is stationary for $\widehat{\mathrm{F}}_{p,q}$ and $q > 0$. Since $\mu(?) = \widehat{\mathrm{F}}_{p,q}\,\mu(?) = r\mu(\widehat{***})$ from (4.4), we conclude that $\mu(?) = 0$ when μ is stationary for $\widehat{\mathrm{F}}_{p,q}$ and $q > 0$.

We now come to the analysis when $q = 0$, which forces $p > 0$ from the definition of $\mathcal{S}$ in (1.1).

Lemma 4.4. *When $q = 0$, we obtain $\mu(?) = 0$ from* (4.70) *for any stationary distribution μ of $\widehat{\mathrm{F}}_{p,q}$.*

Proof. When $q = 0$, forcing $p > 0$, the coefficients $-[p(1-r)+q]$ of $\mu(10?)$ and $-2p^2r^3(1-p)(1+p-q)$ of $\mu(000?1)$ in (4.70) are both strictly negative, which implies that $\mu(10?) = \mu(000?1) = 0$. The former, via (4.29), yields $\mu(000\widehat{**}) = 0$, and this, together with the latter, yields

$$\mu(000?) = \mu(000??) + \mu(000?0) + \mu(000?1) = 0.$$

We now focus on finding $\widehat{\mathrm{F}}_{p,q}\,\mu(000?)$.

In order for $(\widehat{\mathrm{F}}_{p,q}\,\eta(0), \widehat{\mathrm{F}}_{p,q}\,\eta(1), \widehat{\mathrm{F}}_{p,q}\,\eta(2), \widehat{\mathrm{F}}_{p,q}\,\eta(3))$ to equal $(000?)$, we must have $(\eta(3), \eta(4), \eta(5)) \in \widehat{***}$. If each of $\eta(0), \eta(1), \eta(2)$ belongs to $\{0, ?\}$, then each of the events $\widehat{\mathrm{F}}_{p,q}\,\eta(i) = 0$, for $i = 0, 1, 2$, happens with probability p. If $\eta(0) = 1$ and $\eta(1), \eta(2) \in \{0, ?\}$, then $\widehat{\mathrm{F}}_{p,q}\,\eta(0) = 0$ happens with probability $1-q$ and each of $\widehat{\mathrm{F}}_{p,q}\,\eta(1) = 0$ and $\widehat{\mathrm{F}}_{p,q}\,\eta(2) = 0$ happens with probability p. If $\eta(1) = 1$ and $\eta(2) \in \{0, ?\}$, then each of $\widehat{\mathrm{F}}_{p,q}\,\eta(0) = 0$ and $\widehat{\mathrm{F}}_{p,q}\,\eta(1) = 0$ happens with probability $1 - q$ and $\widehat{\mathrm{F}}_{p,q}\,\eta(2) = 0$ happens with probability p. If $\eta(2) = 1$, then each $\widehat{\mathrm{F}}_{p,q}\,\eta(i) = 0$, for $i = 0, 1, 2$, happens with probability $1 - q$. Combining everything, we have

$$\widehat{\mathrm{F}}_{p,q}\,\mu(000?) = p^3 r\mu(\{0,?\}^3\widehat{***}) + (1-q)p^2 r\mu(1\{0,?\}^2\widehat{***}) + (1-q)^2 pr\mu(1\{0,?\}\widehat{***}) + (1-q)^3 r\mu(1\widehat{***}).$$

When $\mu(000?) = 0$, $q = 0$ (and hence $p > 0$), and μ is stationary for $\widehat{\mathrm{F}}_{p,q}$, we conclude from the previous paragraph that $\mu(\{0,?\}^3\widehat{***}) = \mu(1\{0,?\}^2\widehat{***}) = \mu(1\{0,?\}\widehat{***}) = \mu(1\widehat{***}) = 0$. Adding these, we get $\mu(\widehat{***}) = 0$, and via (4.4), this, once again, implies that $\mu(?) = 0$. ☐

5 A formal game theoretic formulation of the problem

We provide a slightly more general game theoretic formulation of our games to facilitate the subsequent discussion on the connection of our games with various other games in the literature. Furthermore, we discuss some interesting open problems in this area. To this end, we shall consider $N = \{1, \ldots, n\}$ to be the (finite) set of players participating in the game, with each player $i \in N$ allowed to choose from a (finite) set of *actions* denoted by A_i. Let $T = \mathbb{N}$ be the set of *rounds* in the game, and let $T_0 = T \cup \{0\}$. A *mover-sequence* is an (infinite) sequence $\mu : T \to N$, such that $\mu(t)$, for $t \in T$, indicates the player who is supposed to make their move in round t of the game.

Given a mover-sequence μ, a *move-sequence* (also referred to as a *history*) is a (finite or infinite) sequence of actions $\underset{\sim}{a} = (a_1, a_2, \ldots)$ such that $a_k \in A_{\mu(k)}$ for each k. We denote by $\mathcal{A}_k$ the set of all move-sequences of length k, for each $k \in T_0$ (note that the move-sequence of length 0 is referred to as the *empty move-sequence* and denoted by $\underset{\sim}{a}^0$). We let $\mathcal{A}_\infty$ denote the set of all move-sequences of infinite length, and $\mathcal{A}$ denote the set of *all* move-sequences, both finite and infinite.

Each move-sequence $\underset{\sim}{a} \in \mathcal{A}$ is assigned a random variable $X_{\underset{\sim}{a}}$ taking values in $\mathbb{R}^n$, which denotes the (random) utilities of the players corresponding to the move-sequence $\underset{\sim}{a}$. The game starts once a realization $(x_{\underset{\sim}{a}})_{\underset{\sim}{a} \in \mathcal{A}}$ of values of the collection of random variables $(X_{\underset{\sim}{a}})_{\underset{\sim}{a} \in \mathcal{A}}$ has been fixed and revealed to all the players.[1] A *strategy* for player i is a collection of functions $\sigma_i := \{\sigma_i^k : k \in T \text{ such that } \mu(k) = i\}$, where $\sigma_i^k : \mathcal{A}_{k-1} \to A_i$. A collection of strategies $\sigma_N := (\sigma_1, \ldots, \sigma_n)$ is called a *strategy profile*.

Each strategy profile σ_N induces an infinite sequence of actions $\underset{\sim}{a}(\sigma_N) := (a_1, a_2, \ldots)$ in the following, natural way: $a_1 = \sigma^1_{\mu(1)}$, $a_2 = \sigma^2_{\mu(2)}(a_1))$, $a_3 = \sigma^3_{\mu(3)}(a_1, a_2)$, and so on. For $t \in T$, we denote by $\underset{\sim}{a}^t(\sigma_N)$ the (truncated) sequence $(a_1, \ldots, a_t)$ of $\underset{\sim}{a}(\sigma_N)$ that is the *game path* of length t induced by σ_N. The *stage utility* of player i at time t, $t \in T$, is given by $u_i^t(\sigma_N) := x_{\underset{\sim}{a}^t(\sigma_N)}(i)$. The *final utility* $u_i(\sigma_N)$ of player i

[1] It is worth emphasizing here that these aren't games with incomplete information as there is no randomness involved once the game begins.

corresponding to a strategy profile σ_N is a (real-valued) function of the stage utility sequence $(\underset{i}{\underset{\sim}{u}}^t(\sigma_N))_{t\in T}$.

We assume, in line with game theory, that each player $i \in N$ plays the game *optimally*. In other words, given a collection of strategies $\sigma_{N\setminus\{i\}}$ of the other players, i plays a *best response*: a strategy σ_i^* that maximizes her utility given $\sigma_{N\setminus\{i\}}$, i.e. $u_i(\sigma_i^*, \sigma_{N\setminus\{i\}}) \geqslant u_i(\sigma_i, \sigma_{N\setminus\{i\}})$ for all strategies σ_i of player i. A strategy σ_i for player i is *dominant* if it is a best response to *every* collection of strategies of the other players. In other words, no matter what strategies the other players adopt, playing σ_i is the best for player i, i.e. $u_i(\sigma_i, \sigma_{N\setminus\{i\}}) \geqslant u_i(\sigma_i', \sigma_{N\setminus\{i\}})$ for all strategies σ_i' of player i and all collections of strategies $\sigma_{N\setminus\{i\}}$ of the other players.[2] A strategy profile σ_N is said to be a *Nash equilibrium* if it is a best response to itself, i.e. for each $i \in N$, σ_i is a best response to $\sigma_{N\setminus\{i\}}$.

Several interesting games can be constructed by imposing restrictions and dependence structures on the random variables $(X_{\underset{\sim}{a}})_{\underset{\sim}{a}\in\mathcal{A}}$ (and accordingly, on the strategy-profiles). A class of such games are the ones where there are two players making alternative moves (that is, $\mu = (1, 2, 1, 2, \ldots)$) and having the same set of actions (i.e. $A_1 = A_2$). For such games, there is an equivalence relation $\sim$ on $\mathcal{A}$ such that $X_{\underset{\sim}{a}^t} = X_{\underset{\sim}{b}^t}$ and $\sigma^t_{\mu(t)}(\underset{\sim}{a}^t) = \sigma^t_{\mu(t)}(\underset{\sim}{b}^t)$ whenever $\underset{\sim}{a}^t$ and $\underset{\sim}{b}^t$ are equivalent.

Combinatorial games played on directed graphs can be modeled via this approach. Let G be a directed graph such that each vertex of G has the same number of outgoing edges, and the outgoing edges from each vertex are identified with the elements of the set A of common actions (where $A = A_1 = A_2$).[3] Then, given an initial vertex, every strategy profile σ_N induces a (directed) path in the graph G. We call two game paths $\underset{\sim}{a}^t$ and $\underset{\sim}{b}^t$ of length t equivalent if they lead to the same vertex in G. Suppose that the utilities are zero-sum, that is, the sum of the utilities of all the players corresponding to any realization of $X_{\underset{\sim}{a}^t}$ is zero. Let the final utility $u_i(\sigma_N)$ of any player $i \in N$ be the first non-zero value in the sequence of stage utilities $(u_i^t(\sigma_N))_{(t\in T_0)}$. In other words, the game ends whenever one of the players receive a non-zero stage utility and the utilities corresponding to that stage become the final utilities. We say that the game is *winning* for player i, for any $i \in N$, if player i has a dominant strategy which, when adopted, rewards her 1 as her final utility. The game ends in a draw if neither player has a winning dominant strategy.

Suppose, further, that $X_{\underset{\sim}{a}^t}(\mu(t))$ equals 1 with probability p, equals -1 with probability q, and equals 0 with probability $1-p-q$, for all $t \in T_0$ (here, we follow the convention that $\mu(0) = 2$). This leads to an (i.i.d.) percolation game on a k-regular tree when $|A| = k$ and each equivalence class is a singleton.[4] On the other hand, fixing an initial vertex arbitrarily (say, the origin), defining $A = \{(0,1),(1,1),(2,1)\}$ (respectively, $A = \{(0,1),(1,1)\}$), and defining two game paths $\underset{\sim}{a}^t = (a^1, \ldots, a^t)$ and $\underset{\sim}{b}^t = (b^1, \ldots, b^t)$ to

[2] It is well-known that a dominant strategy for a player may not exist unless the game has some particular structure.

[3] One can model the games played on arbitrary directed graphs by (recursively) defining the set of actions A_i^t (for player i who is supposed to make their move in round t) as a function defined on the set of possible histories of length $t-1$. In fact, in addition to the utilities and the set of actions, which player is allowed to make a move in a given round can also be a function of the histories. In an even more general setup, all these three variables can be defined randomly corresponding to every history. See [16] (and §6.2 where we discuss) for a formal definition of such random games when the corresponding graph is a tree and the said variables are i.i.d over histories.

[4] See [18] for a formal definition of iid percolation games.

be equivalent if $\sum_{s=1}^{t} a^s = \sum_{s=1}^{t} b^s$, one obtains the percolation game we have studied in this paper (respectively, the percolation game considered in [25]).

Clearly, under such an optimality assumption, a player will play a dominant strategy whenever she has one. We explore what happens to the flow of the game if none of the players has a dominant strategy, limiting ourselves to the case where $|N| = 2$. Suppose player 1 plays a strategy σ_1. Since σ_1 is not dominant, there is a strategy σ_2 of player 2 which, if adopted by player 2, prevents player 1 from winning. Suppose player 2 wins corresponding to the strategy-profile (σ_1, σ_2). However, since σ_2 is not dominant, player 1 has a strategy, say σ_1', such that player 2 does not win when player 1 adopts σ_1'. Continuing in this manner, it follows that when both players play optimally (in other words, under any Nash equilibrium), neither player wins the game, allowing the game to continue forever with each player receiving zero utility at every stage of the game. Summarizing, under the assumption of optimal play, a draw implies continuation of the game for infinite time, which, in addition to requiring the existence of an infinite open path in the (directed) graph, requires that the path be *self-enforcing*, i.e. that neither of the players has an incentive to deviate from this path. The event of an infinite path in a random graph is a prime object of interest in percolation theory, and consequently, the event of a draw is an important object of study in percolation games.

6 Relation with existing literature and open problems

Researchers mainly explore two types of problems in the context of combinatorial games played in random environments, the first of which is concerned with the probability of draw, while the second type focuses on the value of the game.

6.1 Discussion on draw probabilities

As explained in the previous section, the probability of draw, especially whether it is always (i.e. irrespective of the values of the parameters p and q) 0 or not, is an interesting question in percolation games. Apart from [25] and this paper, we are not aware of any work, so far, that looks into this problem.

6.1.1 Generalizing the set of actions

For problems related to the probability of draw in percolation games on $\mathbb{Z}^2$ (or even on arbitrary $\mathbb{Z}^k$), the set of actions A can be generalized in three ways. The first of these is achieved by adding more vertices at the 'next' level, where the 'next' level refers to the value of the y-coordinate in the set A of actions being equal to 1 (for instance, $A = \{(0, 1), \dots, (k, 1)\}$ for some $k \in \mathbb{N}$). The second possibility is the addition of vertices at levels that are 'higher' than the 'next' level, i.e. in which the value of the y-coordinate of at least one of the vertices in the set A of actions is strictly larger than 1 (for instance, $A = \{(0, 1), (1, 2), (2, 1)\}$). The third option is to make A a (simple and structured) function $A(x, y)$ of $(x, y) \in \mathbb{Z}^2$. Such an instance has been dealt with in [5], and the percolation game considered therein has been shown to arise when edge-percolation is considered (i.e., instead of the vertices, it is the edges that are labeled, independently, as trap, target or open with probabilities p, q and $1-p-q$ respectively). The 'location dependent' action set $A(x, y)$ for the edge-percolation game studied in [5] depends (only) on the parity of the first coordinate of (x, y) (it is shown in [5] that $A(x, y) = \{(0, 1), (1, 1)\}$ when x is even, and $A(x, y) = \{(1, 1), (1, 2)\}$ when x is odd).

6.1.2 Arbitrary mover-sequence

Both [25] and this paper consider an alternating mover-sequence (i.e., $\mu = (1, 2, 1, 2 \ldots)$). We would like to explore what happens (in particular, whether the probability of draw continues to be 0 or not) for other mover-sequences (that have been endowed with enough structure, such as when $\mu = (1, 1, 2, 2, 1, 1, 2, 2, \ldots)$). Technically, this relates to a PCA whose update rule is dependent on time. To the best of our knowledge, ergodicity properties of such PCAs have not been explored in the literature, and we think it is an important problem as it generalizes the commonly studied notion of PCAs to a much broader perspective.

6.1.3 Generalizing the event of draw

Yet another class of open problems is concerned with generalizing the events of winning / losing / drawing. Let $u_N^t = (u_1^t, \ldots, u_n^t) \in \mathbb{R}^n$ denote a utility vector for all players in the stage game at time t (the *stage game* refers to the base game, so that the actual game is made up of repetitions of the stage game, and after each stage game is over, every player receives some utility). Given a set of sequences of utility vectors $W_i \subseteq \{(u_N^t)_{t \in T_0}\}$, let us call a game *winning* for player i if player i has a dominant strategy to ensure a utility sequence in W_i, and a game is said to *lead to a draw* if it is not winning for any of the players. A natural extension of the existing notion of winning (as defined in our paper and in [25]) is obtained by defining W_i as the set of utility sequences for which i receives a cumulative utility of amount k, for some pre-fixed $k \in \mathbb{N}$, before any other player does. More formally, W_i contains those utility-vector sequences $(u_N^t)_{t \in T_0}$ for which there is $\hat{t} \in T_0$ such that $\sum_{t=0}^{\hat{t}} u_i^t = k$ and $\sum_{t=0}^{\bar{t}} u_j^t < k$ for all $j \in N \setminus \{i\}$ and all $\bar{t} < \hat{t}$. The existing notion of winning corresponds to the situation where $k = 1$. We think it is an interesting problem to explore what happens for higher values of k.

6.1.4 Percolation games on three (or higher) dimensions

The set-up for percolation games can be generalized to lattices in any dimension $k \in \mathbb{N}$. Each site $\mathbf{x} = (x_1, x_2, \ldots, x_k) \in \mathbb{Z}^k$ is assigned, independently, the label of trap with probability p, the label of target with probability q, and the label of open with probability $1 - p - q$, and A can be defined as a subset of $\mathbb{Z}^k$. An even more general set-up, on arbitrary, locally finite directed graphs is addressed in §1.2 of [25], but with $q = 0$ (i.e. there are no target-labeled vertices, and hence such a percolation game is referred to as a *trapping game*). The *dimension reduction method* is illustrated in [25], showing that if the directed graph under consideration satisfies certain conditions, then the existence of multiple Gibbs states for the hard-core model on a related graph would imply that the trapping game on the directed graph has a positive probability of culminating in a draw.

It is worthwhile to ponder if there are regime(s), as function(s) of the parameters p and q, in which percolation games played on lattices of higher dimensions have probability 0 of resulting in a draw, and whether, when A is suitably defined, the occurrence of draw in such a game can be connected with the ergodicity of a PCA, possibly deduced via the recurrence relations arising from the game, the same way as illustrated in this paper. An immediate problem to consider can be where we set $A = \{(0, 0, 1), (1, 0, 1), (0, 1, 1)\}$ and the same equivalence relation as ours (that is, $(a_1, \ldots, a_k) \sim (b_1, \ldots, b_l)$ if $\sum_{j=1}^k a_j = \sum_{j=1}^l b_l$). This leads to a nice class of percolation games on $\mathbb{Z}^3$. Attempting to implement the technique of weight functions to find conditions on p and q that guarantee ergodicity of the PCA deduced from the recurrence relations of this game would be an interesting problem for the near future.

6.1.5 Monotonicity of draw probability

An important problem, in our opinion, of a slightly different flavour, is the analysis of monotonicity properties, if any, of the probabilities of draw in percolation games with respect to, in some sense, the degree of '*mixing*' of the actions / moves. Consider a class of percolation games on $\mathbb{Z}^2$, where each site (x, y) is assigned the label $X_{(x,y)}$, with $X_{(x,y)}$ i.i.d. over all $(x, y) \in \mathbb{Z}^2$ and $X_{(x,y)}$ equaling trap with probability p, target with probability q, and open with probability $1 - p - q$. We characterize the games we are concerned with in this discussion by two parameters: k and ℓ, where $k \geqslant 2$ is the number of actions allowed for each player, and $1 \leqslant \ell \leqslant k$ is a mixing parameter. For every position of the game (x, y) in $\mathbb{Z}^2$ (i.e. (x, y) is where the token is currently located), the action a_i leads to the site $((x - 1)\ell + i, y + 1)$, for all $1 \leqslant i \leqslant k$. In other words, here, the set of outgoing edges $\mathrm{Out}(x, y)$ from (x, y) is $\{((x - 1)\ell + i, y + 1) : 1 \leqslant i \leqslant k\}$.

Note that, for any two '*neighboring*' game positions (x, y) and $(x+1, y)$, the cardinality of $\mathrm{Out}(x, y) \cap \mathrm{Out}(x + 1, y)$ is precisely $k - \ell$. In particular, given k, when $\ell = 1$, we obtain the k-neighbor percolation game on the integer lattice $\mathbb{Z}^2$ (for instance, [25] considers $k = 2$, and this paper is concerned with $k = 3$), and when $\ell = k$, we obtain the percolation game on a rooted k-regular tree. It has been shown in [28] that for arbitrary k, when $\ell = k$, there are positive values of p and / or q for which the probability of draw is strictly positive in the percolation game played on a rooted k-regular tree (more precisely, the probability of draw is strictly positive whenever $(1 - p - q)(1 - q)^{k-1} \leqslant (k + 1)^{k-1}k^{-k}$). At the other extreme, for $k = 2, 3$ and $\ell = 1$, it has been shown in [25] and this paper that both p and q need to equal 0 for the probability of draw in the corresponding percolation game to be strictly positive. These findings raise the following important questions: does draw become less likely

1. as ℓ decreases (i.e. there is a greater degree of mixing) for a given k,
2. as k increases while ℓ remains fixed (for instance, when ℓ remains fixed at the value $\ell = 1$).

It is worth emphasizing that for a given, arbitrary configuration of trap / target / open labels (i.e. a given realization of the i.i.d. random variables $(X_{(x,y)})_{(x,y)\in\mathbb{Z}^2}$), no monotonicity can be found for the event of draw with respect to ℓ in the two situations mentioned above, implying that a coupling argument will not be sufficient to prove such results. We believe that a more involved combinatorial argument will be needed for addressing these questions, and currently, we are pursuing our research exploring these questions. Exploring such monotonicity questions for just the *trapping game* (where $q = 0$) or just the *target game* (where $p = 0$) is expected to be quite challenging.

6.2 Discussion on the value of a game

The *value* of a zero-sum game is a well-known quantity in game theory, but considering readers from various backgrounds, we provide its definition here. The value of a zero-sum game is said to exist when the utilities for player 1, obtained via the following two 'ways of thinking', turn out to be equal to each other, and the value is defined as the common utility (whenever it exists). In the first way of thinking, player 1 believes that player 2's only intention is to 'hurt' her, that is, whatever strategy σ_1 she plays, player 2 will play a strategy σ_2 that minimizes her utility given σ_1. Let $U^{\max\min}$ be the maximum utility player 1 can guarantee given this belief. Formally, $U^{\max\min} = \max_{\sigma_1\in\Sigma_1} \min_{\sigma_2\in\Sigma_2} u_1(\sigma_1, \sigma_2)$, where Σ_1 and Σ_2 are the sets of all possible strategies for players 1 and 2 respectively. In the second way of thinking, whatever strategy σ_2 is played by player 2, player 1 plans to play a strategy σ_1 that maximizes her utility given σ_2, and believes that player 2 will play the strategy that minimizes her utility when she plays according to this plan. Let

$U^{\min\max} = \min_{\sigma_2 \in \Sigma_2} \max_{\sigma_1 \in \Sigma_1} u_1(\sigma_1, \sigma_2)$ be the utility player 1 receives according to this belief. The value of a game exists when $U^{\min\max} = U^{\max\min}$, and this common utility is called the value of the game.

In [18], a general notion of zero-sum percolation games on $\mathbb{Z}^d$ is defined, with two players making alternating moves.[5] In contrast to our notion of stage games where exactly one player makes a move in each round, [18] defines a stage game in which each round comprises a move by player 1 followed immediately by a move by player 2. Each of the players receives some utility at the end of each stage game. An n-stage game, for $n \in \mathbb{N}$, consists of n stage games played sequentially where the utility of each player is defined as the time-average (where time refers to the number of stages) of the total utility from all the stage games of that player. The utility function of an n-stage game can be defined using our framework by requiring that $u_i^t(\sigma_N) = 0$ for all $t \in 2\mathbb{N} - 1$ and all $t > n$, and defining $u_i(\sigma_N) = \frac{1}{n} \sum_{t \in T} u_i^t(\sigma_N)$.

For i.i.d. and oriented percolation games, [18] shows that as n tends to infinity, the sequence of (random) values of n-stage games converges almost surely to a constant, and the expected value converges at a rate of $O(\ln(n) n^{-1/2})$. They further show that the assumption of i.i.d. is necessary for this result. It is worth mentioning that the games we have considered in our paper are i.i.d. and oriented. We refer the reader to [18] for a formal definition of i.i.d. and oriented percolation games.

In [16], a general class of two-player zero-sum games is considered where, in addition to the utilities (that are defined as the capacities corresponding to any history), in every round of the game, the mover and the set of available actions are both random and i.i.d. over the histories. The three random variables, i.e. the chosen mover, the chosen set of available actions, and the utilities, need not be independent of each other corresponding to a given history. The games in [16] are played on (finitely branching) trees for each possible realization of the above-mentioned random variables. The utility for player 1 corresponding to a strategy profile is the infimum of the stage utilities along the game path induced by the strategy profile. In this setting, a natural measure space is defined over the games, and a (fixed point) characterization of the cumulative distribution function for the value of the game is provided (together with a number of corollaries and applications).

Although [2] does not deal with the value of a game nor its probability of draw, uses an approach involving ideas from percolation theory which makes it relevant for percolation games. In [2], random games are considered where the number of players is large, each player has two strategies, and the utilities are i.i.d. with possible ties (i.e. having an atomic distribution). In this setting, [2] explores the possibility of existence of a (pure) Nash equilibrium which, in contrast with mixed Nash equilibrium, may not always exist. In particular, asymptotic results regarding the (random) number of Nash equilibria and a central limit theorem for given bounds on the probability of ties are established. Furthermore, using tools from percolation theory, the geometry of the set of Nash equilibria is determined, and it is shown that the best response dynamics converges to a Nash equilibrium with high probability when the number of players is large and the probability of ties is positive but bounded above (such as in the case of potential games).[6]

[5] Informally speaking, a percolation game requires, among other criteria, the (random) utility functions to be stationary and ergodic.

[6] More precisely, [2] establishes a connection between percolation and random oriented graphs through a coupling where the set of strategies that are accessible by the best response dynamics coincides with the connected component containing the origin for percolation on the hypercube.

7 The scope of the weight function technique

Let us define the critical region for a percolation game as the set of all values of $(p,q) \in \mathcal{S}$ (where $\mathcal{S}$ is as defined in (1.1)) for which the probability of draw is 0 (and hence, the corresponding PCA is ergodic). It follows from this paper and [25] that the critical region for the percolation games considered in both these papers is $\mathcal{S}$.

We begin with noting that the weight function technique provides only a sub-region, normally in terms of a lower bound, of the critical region of a percolation game. In case of our percolation games or that in [25], this sub-region turns out to be 'full', i.e. it covers all possible values of (p,q), and consequently, the weight function technique succeeds in providing a characterization of the *entire* critical region. In general, unless the critical region is full in the above-mentioned sense, the weight function technique itself will not be suffice to characterize the critical region, which, in our view, is the main weakness of this technique. Nevertheless, providing a sub-region of the critical region is, in general, an important contribution to the literature (in fact, it may well be the case that finding such a sub-region becomes the first breakthrough in answering questions about the critical regions in several open problems), and in view of that, we now proceed to describe some situations where one can apply this technique.

Recall that the weight function technique, when applicable, only shows that certain events have probability 0 under *any* stationary distribution corresponding to the PCA under consideration. One then needs to do further computations to deduce how this actually leads to the conclusion that the event of draw has probability 0 under *any* stationary distribution. The crucial step here is to link the ergodicity of the PCA under consideration with the event of draw, which may not be straightforward unless the PCA has some game theoretic structure. In what follows, we define a class of 'game theoretic' PCAs for which we believe that the weight function technique can be applied to establish ergodicity. For PCAs outside of this class, our guess is that the applicability of the weight function technique will depend on the particular structure of the PCA's stochastic update rule, but as of now, we are unable to specify what structural conditions on the update rule guarantee such applicability.

We call a d-dimensional deterministic CA F, with universe $\mathbb{Z}^d$ and a given neighborhood $\mathcal{N}$ that is a finite subset of $\mathbb{Z}^d$, *game-theoretic* if it satisfies the following criteria:

1. the alphabet $\widehat{\mathcal{A}}$ comprises the symbols W, D and L;
2. letting $\eta_t = \left(\eta_t(\mathbf{x}) : \mathbf{x} \in \mathbb{Z}^d\right)$ denote the configuration of states at time t, where $\eta_t(\mathbf{x})$ denotes the state of $\mathbf{x}$ at time t, the state $\eta_{t+1}(\mathbf{x})$ of *any* site $\mathbf{x} \in \mathbb{Z}^d$ at time $t+1$ is decided according to the following rules:
   (a) if $\eta_t(\mathbf{x}+\mathbf{y}) = W$ for *each* $\mathbf{y} \in \mathcal{N}$, then $\eta_{t+1}(\mathbf{x}) = L$,
   (b) if there exists at least one $\mathbf{y} \in \mathcal{N}$ for which $\eta_t(\mathbf{x}+\mathbf{y}) = L$, then $\eta_{t+1}(\mathbf{x}) = W$,
   (c) if there exists *no* $\mathbf{y} \in \mathcal{N}$ for which $\eta_t(\mathbf{x}+\mathbf{y}) = L$, but there exists at least one $\mathbf{z} \in \mathcal{N}$ such that $\eta_t(\mathbf{x}+\mathbf{z}) = D$, then $\eta_{t+1}(\mathbf{x}) = D$;
3. conditioned on η_t, the updates $\eta_{t+1}(\mathbf{x})$ happen independently over all $\mathbf{x} \in \mathbb{Z}^d$.

We obtain a *game-theoretic PCA* by perturbing the above CA in a manner similar to what we have considered in this paper: given parameters p and q, each outcome (of the above-mentioned deterministic CA) that is not equal to W is flipped to W with proability p, and each outcome that is not equal to L is flipped to L with probability q. More formally, letting $\hat{\eta}_t = \left(\hat{\eta}_t(\mathbf{x}) : \mathbf{x} \in \mathbb{Z}^d\right)$ denote the (random) configuration of states at time t,

1. we let $\hat{\eta}_{t+1}(\mathbf{x})$ equal L with probability $1-p$ and W with probability p when $\eta_{t+1}(\mathbf{x}) = L$;

2. we let $\hat{\eta}_{t+1}(\mathbf{x})$ equal W with probability $1-q$ and L with probability q when $\eta_{t+1}(\mathbf{x}) = W$;
3. we let $\hat{\eta}_{t+1}(\mathbf{x})$ equal W with probability p, L with probability q and D with probability $1-p-q$ when $\eta_{t+1}(\mathbf{x}) = D$.

We observe, crucially, that for game-theoretic PCAs, both parts of Lemma 3.1 hold (and this can be proved in an identical manner as shown in this paper). The first part of Lemma 3.1 leads to the conclusion, via a proof identical to that of Proposition 2.1 of [25], that F is ergodic if and only if the PCA A, obtained by restricting the alphabet $\hat{\mathcal{A}}$ to the sub-alphabet $\mathcal{A} = \{W, L\}$, is ergodic. Moreover, the unique stationary or limiting distribution μ of F, when F is ergodic, is the same as that of A, and therefore, μ must assign probability 0 to the symbol D, i.e. the probability, under μ, of the event that the symbol D occupies the site $\mathbf{x}$ is 0 for *every* $\mathbf{x} \in \mathbb{Z}^d$. Applying the technique of weight functions to the PCA F, provided the process of constructing the weight function is tractable either manually or with the aid of a computer, can reveal conditions on the parameter pair (p, q) under which *any* stationary distribution μ for F assigns probability 0 to the symbol D.

We believe that the method of weight function may even be of aid in case of percolation games (on lattices) whose recurrence relations do not necessarily yield a PCA, but something more general. For instance, suppose we consider the same random premise as in this paper, i.e. each site of $\mathbb{Z}^2$ is assigned, independently, a label that reads trap with probability p, target with probability q, and open with probability $1-p-q$, but here, for each site $(x, y) \in \mathbb{Z}^2$, we set $\mathrm{Out}(x, y) = \{(x+2, y), (x+2, y+2), (x, y+2)\}$. As before, two players take turns to make moves, where a move involves relocating the token from where it is currently located, say some site (x, y), to any of the sites in $\mathrm{Out}(x, y)$. The outcome of this game is decided following the same rules as the percolation game that we study in this paper. This problem has also been mentioned in §6.1.1, when discussing the possibility of adding vertices at levels that are higher than the next level.

It is immediate that the recurrence relations arising from the above game cannot be represented via a PCA, but could be by a PCA of memory two as defined in [11]. A little pondering reveals that the only way to partition the lattice $\mathbb{Z}^2$ into "level sets" that are meaningful for our analysis of this game is to consider the diagonal lines $D_k = \{(x, k-x) : x \in \mathbb{Z}\}$, identify D_k with $\mathbb{Z}$ via the mapping $(x, k-x) \mapsto x$ for each $x \in \mathbb{Z}$, and consider the following discrete-time stochastic process $\{\eta_t\}_{t \in \mathbb{N}_0}$ where each η_t is a random configuration taking values in $\hat{\mathcal{A}}^{\mathbb{Z}} = \{W, D, L\}^{\mathbb{Z}}$. Conditioned on η_{t-1} and η_t, the random variables $\eta_{t+1}(x)$ are defined independently over all $x \in \mathbb{Z}$, with the distribution for $\eta_{t+1}(x)$ being a function of $\eta_t(x)$, $\eta_{t-1}(x+2)$ and $\eta_t(x+2)$, as follows:

1. if $\eta_t(x) = \eta_{t-1}(x+2) = \eta_t(x+2) = W$, then $\eta_{t+1}(x) = L$ with probability $1-p$ and $\eta_{t+1}(x) = W$ with probability p,
2. if at least one of $\eta_t(x)$, $\eta_{t-1}(x+2)$ and $\eta_t(x+2)$ equals L, we set $\eta_{t+1}(x) = W$ with probability $1-q$ and $\eta_{t+1}(x) = L$ with probability q,
3. if none of $\eta_t(x)$, $\eta_{t-1}(x+2)$ and $\eta_t(x+2)$ equals L but at least one of them equals D, we set $\eta_{t+1}(x) = W$ with probability p, $\eta_{t+1}(x) = L$ with probability q, and $\eta_{t+1}(x) = D$ with probability $1-p-q$.

It is evident that $\{\eta_t\}_{t \in \mathbb{N}_0}$ does not constitute a Markov chain. However, it is Markov of order 2 (i.e. retains as memory the configurations of the previous 2 time steps), and a probability distribution μ supported on $\hat{\mathcal{A}}^{\mathbb{Z}} \times \hat{\mathcal{A}}^{\mathbb{Z}}$ is said to be stationary or invariant for this stochastic process if the following is true: if the joint distribution of (η_{t-1}, η_t) is μ, then the joint distribution of (η_t, η_{t+1}) is μ as well. We then say that $\{\eta_t\}_{t \in \mathbb{N}_0}$ is ergodic if

1. it possesses a unique stationary distribution μ, and

2. if ν denotes *any* probability distribution supported on $\hat{\mathcal{A}}^{\mathbb{Z}} \times \hat{\mathcal{A}}^{\mathbb{Z}}$, and if the joint distribution of (η_0, η_1) is ν, then the joint distribution of (η_t, η_{t+1}) converges to μ as $t \to \infty$.

It should be possible to prove, along the same lines as Proposition 2.2 of [25] has been argued, that the percolation game described above has probability 0 of culminating in a draw if and only if the process $\{\eta_t\}_{t\in\mathbb{N}_0}$ is ergodic. Once again, it becomes important to investigate the critical region for this percolation game. We expect that the technique of weight functions is applicable in this set-up. Here, we must look for a suitable function w that is defined on the space of probability measures supported on $\hat{\mathcal{A}}^{\mathbb{Z}} \times \hat{\mathcal{A}}^{\mathbb{Z}}$, and we suspect that we have to consider measures of Carteisan products of cylinder sets (in other words, we have to consider events of the form $(\eta_{t-1}(x) = a_x$ for all $x \in A, \eta_t(y) = b_y$ for all $y \in B)$, where A and B are finite subsets of $\mathbb{Z}$, and $a_x, b_y \in \hat{\mathcal{A}}$ for all $x \in A$ and $y \in B$).

References

[1] Marie Albenque, *A note on the enumeration of directed animals via gas considerations*, The Annals of Applied Probability **19** (2009), no. 5, 1860–1879. MR2569810

[2] Ben Amiet, Andrea Collevecchio, Marco Scarsini, and Ziwen Zhong, *Pure Nash equilibria and best-response dynamics in random games*, Mathematics of Operations Research **46** (2021), no. 4, 1552–1572. MR4370838

[3] Paul Balister, Béla Bollobás, and Robert Kozma, *Large deviations for mean field models of probabilistic cellular automata*, Random Structures & Algorithms **29** (2006), no. 3, 399–415. MR2254497

[4] Riddhipratim Basu, Alexander E Holroyd, James B Martin, and Johan Wästlund, *Trapping games on random boards*, The Annals of Applied Probability **26** (2016), no. 6, 3727–3753. MR3582816

[5] Dhruv Bhasin, Sayar Karmakar, Moumanti Podder, and Souvik Roy, *Ergodicity of a generalized probabilistic cellular automaton with parity-based neighbourhoods*, arXiv preprint arXiv:2212.01753 (2022).

[6] Yvan Le Borgne and Jean-François Marckert, *Directed animals and gas models revisited*, The Electronic Journal of Combinatorics **14** (2007), no. 1, R71. MR2365970

[7] Olivier Bouré, Nazim Fates, and Vincent Chevrier, *Probing robustness of cellular automata through variations of asynchronous updating*, Natural Computing **11** (2012), no. 4, 553–564. MR3002413

[8] Mireille Bousquet-Mélou, *New enumerative results on two-dimensional directed animals*, Discrete Mathematics **180** (1998), no. 1-3, 73–106. MR1603701

[9] Guy Bresler, Chenghao Guo, and Yury Polyanskiy, *Linear programs with polynomial coefficients and applications to 1d cellular automata*, arXiv preprint arXiv:2204.06357 (2022).

[10] Ana Bušić, Jean Mairesse, and Irene Marcovici, *Probabilistic cellular automata, invariant measures, and perfect sampling*, Advances in Applied Probability **45** (2013), no. 4, 960–980. MR3161292

[11] Jérôme Casse and Irène Marcovici, *Probabilistic cellular automata with memory two: invariant laws and multidirectional reversibility*, Annales Henri Lebesgue **3** (2020), 501–559. MR4149819

[12] Deepak Dhar, *Exact solution of a directed-site animals-enumeration problem in three dimensions*, Physical Review Letters **51** (1983), no. 10, 853. MR0721768

[13] RL Dobrušin, VI Kriukov, and AL Toom, *Stochastic cellular systems: ergodicity, memory, morphogenesis*, Manchester University Press, 1990.

[14] Jean-Christophe Dubacq, *How to simulate turing machines by invertible one-dimensional cellular automata*, International Journal of Foundations of Computer Science **6** (1995), no. 04, 395–402.

[15] Wesam M Elsayed, Mohammed Elmogy, and B El-Desouky, *Evolutionary behavior of dna sequences analysis using non-uniform probabilistic cellular automata model*, Cincia e Tcnica Vitivincola **32** (2017), 137–148.

[16] János Flesch, Arkadi Predtetchinski, and Ville Suomala, *Random perfect information games*, Mathematics of Operations Research **48** (2023), no. 2, 708–727. MR4588937

[17] Peter Gács, *Reliable cellular automata with self-organization*, Journal of Statistical Physics **103** (2001), no. 1, 45–267. MR1828729

[18] Guillaume Garnier and Bruno Ziliotto, *Percolation games*, Mathematics of Operations Research (2022).

[19] Antoine Georges and Pierre Le Doussal, *From equilibrium spin models to probabilistic cellular automata*, Journal of Statistical Physics **54** (1989), no. 3, 1011–1064. MR0988570

[20] Eric Goles and Martín Matamala, *Uniform simulation of turing machines by cellular automata*, Cellular Automata and Complex Systems (1999), 23–36. MR1672854

[21] G Grinstein, C Jayaprakash, and Yu He, *Statistical mechanics of probabilistic cellular automata*, Physical Review Letters **55** (1985), no. 23, 2527. MR0819681

[22] I Hartarsky and R Szabó, *Generalised oriented site percolation, probabilistic cellular automata and bootstrap percolation*, arXiv preprint (2021). MR4408459

[23] Ivailo Hartarsky, *Bootstrap percolation, probabilistic cellular automata and sharpness*, Journal of Statistical Physics **187** (2022), no. 3, 1–17. MR4408459

[24] Gustav A Hedlund, *Endomorphisms and automorphisms of the shift dynamical system*, Mathematical Systems Theory **3** (1969), no. 4, 320–375. MR0259881

[25] Alexander E Holroyd, Irène Marcovici, and James B Martin, *Percolation games, probabilistic cellular automata, and the hard-core model*, Probability Theory and Related Fields **174** (2019), no. 3, 1187–1217. MR3980314

[26] Jarkko Kari, *Theory of cellular automata: A survey*, Theoretical Computer Science **334** (2005), no. 1-3, 3–33. MR2132942

[27] Jarkko Kari, *Cellular automata*, University of Turku (2013).

[28] Sayar Karmakar, Moumanti Podder, Souvik Roy, and Soumyarup Sadhukhan, *Phase transition in percolation games on rooted galton-watson trees*, arXiv preprint arXiv:2303.11402 (2023).

[29] Lemont B Kier, Chao-Kun Cheng, and Bernard Testa, *A cellular automata model of the percolation process*, Journal of chemical information and computer sciences **39** (1999), no. 2, 326–332.

[30] Joel L Lebowitz, Christian Maes, and Eugene R Speer, *Statistical mechanics of probabilistic cellular automata*, Journal of Statistical Physics **59** (1990), no. 1, 117–170. MR1049965

[31] Qin Lei, Jia Lee, Xin Huang, and Shuji Kawasaki, *Entropy-based classification of elementary cellular automata under asynchronous updating: an experimental study*, Entropy **23** (2021), no. 2, 209. MR4224579

[32] Pierre-Yves Louis and Francesca R Nardi, *Probabilistic cellular automata*, Emergence, Complexity, Computation **27** (2018). MR3791805

[33] Jean Mairesse and Irene Marcovici, *Around probabilistic cellular automata*, Theoretical Computer Science **559** (2014), 42–72. MR3280727

[34] Irène Marcovici, Mathieu Sablik, and Siamak Taati, *Ergodicity of some classes of cellular automata subject to noise*, Electronic Journal of Probability **24** (2019), 1–44. MR3940771

[35] Mark McCann and Nicholas Pippenger, *Fault tolerance in cellular automata at high fault rates*, Journal of Computer and System Sciences **74** (2008), no. 5, 910–918. MR2416359

[36] Mark McCann and Nicholas Pippenger, *Fault tolerance in cellular automata at low fault rates*, Journal of Computer and System Sciences **79** (2013), no. 7, 1126–1143. MR3061405

[37] Kenichi Morita and Masateru Harao, *Computation universality of one-dimensional reversible (injective) cellular automata*, IEICE Transactions (1976–1990) **72** (1989), no. 6, 758–762. MR1178211

[38] E Popova, Y Staraselski, A Brahme, RK Mishra, and K Inal, *Coupled crystal plasticity–probabilistic cellular automata approach to model dynamic recrystallization in magnesium alloys*, International Journal of Plasticity **66** (2015), 85–102.

[39] Damien Regnault, *Directed percolation arising in stochastic cellular automata analysis*, in: International Symposium on Mathematical Foundations of Computer Science, Springer, 2008, pp. 563–574. MR2539401

[40] Pal Rujan, *Cellular automata and statistical mechanical models*, Journal of Statistical Physics **49** (1987), no. 1, 139–222. MR0923854

[41] Ville Salo, Guillaume Theyssier, and Ilkka Törmä, *Bootstrap percolation and cellular automata*, arXiv preprint arXiv:2110.00656 (2021). MR4437739

[42] JW Thatcher, *Self-describing turing machines and self-reproducing cellular automata*, Essays on Cellular Automata (1970), 103–131.

[43] Tommaso Toffoli, *Computation and construction universality of reversible cellular automata*, Journal of Computer and System Sciences **15** (1977), no. 2, 213–231. MR0462816

[44] Vladimir K Vanag, *Study of spatially extended dynamical systems using probabilistic cellular automata*, Physics-Uspekhi **42** (1999), no. 5, 413.

[45] Branislav Vukotić and Milan M Ćirković, *Astrobiological complexity with probabilistic cellular automata*, Origins of Life and Evolution of Biospheres **42** (2012), no. 4, 347–371.

Acknowledgments. We are thankful to the editor and two anonymous referee for their useful feedbacks that certainly helped improve the paper.