Protect the Colours, Rotate the Chords

A Design Architecture for Eight-Person Peer Networks, and an Honest Account of What Is Known About Them

Victor del Rosal
fiveinnolabs / National College of Ireland
Newport, County Mayo, Ireland

August 2026

Abstract

Small deliberate peer groups, variously called masterminds, peer advisory boards or founder circles, are widely sold and, in the material we were able to read, rarely specified. In the material we have seen they are described by their cadence rather than by any mechanism that would explain why they work or predict when they fail. This paper proposes a design architecture for such groups and states its assumptions explicitly enough to be argued with.

We model the group as a set of nodes retaining distinct identities plus a set of edges whose value is multiplicative in three factors: useful difference, trust, and mutual legibility. The multiplicative form predicts failure at both extremes, since a group of diverse strangers and a group of trusting near-clones both produce approximately nothing. It also makes explicit an assumption we did not find stated in the material we read, namely that difference and trust are negatively correlated across edges, so that the most valuable pairings would be the least likely to form spontaneously and the least comfortable when forced early. We assume that correlation rather than demonstrate it, and we know of no direct measurement of it.

We propose resolving this by scheduling rather than exhortation. Placing members on a circle so that angular distance encodes difference, the complete graph on eight vertices admits a one-factorization in which every round is homogeneous in chord length, giving a seven-round schedule that traverses all twenty-eight relationships in non-decreasing order of chord length. Chord length is a proxy for difference and not a measurement of it, so the grading is exact in the schedule and only approximate in the quantity the schedule exists to grade. On our worked example the two come apart at every boundary. We show there are exactly two such factorizations, so the schedule is essentially forced, and we characterise the group sizes admitting one as the powers of two. That characterisation is a corollary of the Stern-Lenz Lemma, and we claim no novelty for any of the three results; only the framing is ours.

We then attend to the assumption the schedule rests on. Throughout, chord distance means the number of seats between two members around the octagon, the graph distance on the cycle, and not the Euclidean length of the chord, which is a concave function of it. Seating members so that chord distance tracks difference is circular seriation, exactly solvable only for circular Robinson matrices. At $n = 8$ the placement problem, NP-hard in general, is solved exactly by enumerating all 2520 arrangements. On a simulated worked example the exact condition fails on every row while the enumeration still recovers the true underlying order, and the residual of the fitted relation is close to the size of one seat step. That residual is a prediction error in dissimilarity units and not a seating error, and its magnitude is set by how strong the second latent dimension is, so the example bounds nothing on its own; we report a sensitivity sweep rather than a single figure.

Finally we situate the design against the field-experimental evidence, which is thinner and less favourable than the promotional material we encountered would suggest. We could find no commercial peer advisory group that has been evaluated with a control group. The randomized evidence that does exist comes from small firms in emerging economies, attributes a substantial part of the gain to peer quality alongside a measured improvement in management practice, and in the nearest available comparison favours stable groups over rotating contact. The paper presents no empirical validation. It is a design proposal, a set of proofs, and a list of ways to be wrong.

Keywords: peer networks, homophily, structural holes, group design, one-factorization, circular seriation, round-robin scheduling, field experiments, collective agency

1 Introduction

The idea that a small group of committed people can make each of its members materially more capable is old, commercially popular, and almost entirely unspecified. Hill's original formulation of the master mind [26] describes a coordination of knowledge and effort in a spirit of harmony, which is evocative and operationally empty. The format is still sold today with much the same description: a curated group, a regular meeting, a facilitator, and a promise of transformation. We have not conducted a systematic review of that material and offer the characterisation as an impression.

The gap between the popularity of the format and the poverty of its specification is the motivation for this paper. It is a wide gap. We could find no peer-reviewed evaluation with a control group of any commercial peer advisory organisation, and such public evidence as we encountered consists of self-commissioned comparisons between members and non-members, which cannot separate treatment from selection. Meanwhile a small number of genuine field experiments on structured peer contact among firm owners do exist, and they are more equivocal than the promotional material we encountered would suggest. Section 2.5 sets both out.

Practitioners readily list four ways these groups fail, and we set them out as the informal claims they are rather than as findings. Convergence, in which members are said to end up deploying the same vocabulary and the same few tactics while homogeneity feels like alignment. Stratification, in which one member becomes the perpetual advice-giver. Softening, in which the accountability step quietly becomes optional. Fragmentation, in which a handful of comfortable relationships thrive while the rest stay cordial. We have no citation for any of these and no data of our own; they are practitioner report, which is the same currency Section 2.5 declines to accept from vendors, and we use them only to motivate a structural description that can then be tested. Each of these failures has a structural description, and structural descriptions admit structural remedies. Our aim is to state the mechanism precisely enough that the remedies can be designed, measured, and disconfirmed.

1.1 Contributions

1. A multiplicative edge-value model with three factors, difference, trust and legibility, and an argument that legibility rather than knowledge transfer is the quantity a peer group primarily accumulates (Section 3).
2. An explicit statement of the difference-trust tension, and the conjecture, labelled as such in Section 4 and unsupported by evidence, that it drives unstructured peer groups onto their least valuable edges (Section 4).
3. A formalisation of the seating problem as circular seriation, with the exact condition under which the octagon embedding exists, an exhaustive placement procedure that is tractable precisely because $n = 8$, and a worked example in which the exact condition fails and the residual is of the same order as one seat (Section 5).
4. A distance-graded one-factorization of K_8, a uniqueness result showing there are exactly two of them, and the characterisation of the sizes admitting such a schedule. We claim novelty for none of the three. The characterisation is a corollary of the Stern-Lenz Lemma; the existence of the K_8 schedule is that corollary evaluated at $2n = 8$; and the uniqueness count follows in a few lines from the textbook fact that an even cycle splits into two perfect matchings. What we believe is new is only the framing, namely ordering the rounds of a round robin by a similarity measure defined on the pairs, and that is a modest contribution. Section 6.4 sets out the relation to prior work in full (Section 6).
5. A four-stage architecture and a single behavioural diagnostic, unprompted routing, proposed as the operational test of whether the group functions (Sections 7 and 8), together with six falsifiable predictions and an honest account of what the existing evidence does to each (Sections 9 and 10).

The paper presents no empirical validation. The mathematics of Sections 5 and 6 is proved and computationally verified; everything else is a design proposal with an argument attached.

2 Related work

The relevant literature is scattered across network sociology, team research, behavioural science and combinatorial design, and to our knowledge has not been assembled into a design specification for groups of this size.

2.1 Difference and its returns

Granovetter's account of weak ties [21] and Burt's work on structural holes [6][7] establish that non-redundant information flows across social distance rather than within cohesive clusters, and that individuals spanning such gaps enjoy measurable advantages. Harrison and Klein's typology of diversity as separation, variety and disparity [24] supplies the distinction the present model needs: the useful axis of difference must be specified rather than assumed, and it is variety rather than demographic separation that the model refers to.

The formal result most often invoked in this area deserves a caveat rather than a citation. Hong and Page [27] prove that under stated conditions a randomly drawn diverse team outperforms a team of the individually ablest, a result popularised as diversity trumping ability [41]. Thompson [52] argues in the Notices of the American Mathematical Society that the theorem is false as stated and close to a restatement of its hypotheses once repaired, and that the accompanying simulation demonstrates the benefit of randomised search rather than of diversity. Page disputes this by appeal to a condition first stated in his 2007 book, and Kuehn [32] defends the original while conceding the point about randomisation. The exchange is unresolved. We therefore treat Hong and Page as a formal model with contested foundations, and we do not rest any claim in this paper on it.

2.2 The cost of difference

The countervailing findings matter more for design. McPherson, Smith-Lovin and Cook document homophily as a pervasive organising force [38]; ties form preferentially between similar people, so the edges that would carry the most novel information are the least likely to exist. Uzzi [53] shows that embeddedness carries both benefits and a penalty at the extremes. Reagans and McEvily [43] find that cohesion and network range independently ease knowledge transfer, implying that distance without cohesion transfers poorly. Aral and Van Alstyne [2] state the trade-off most directly: diverse ties carry more novel information per unit but support lower communication bandwidth, so total novelty received is not monotonically increasing in diversity. The multiplicative model of Section 3 is an attempt to compress these findings into something a group designer can act on.

2.3 What groups do with what they know

Stasser and Titus [49] demonstrate that groups over-discuss information already held in common and under-discuss information held uniquely. This is among the better-replicated findings in the area: a meta-analysis of 65 studies covering 3,189 groups [37] finds groups mentioned shared information roughly two standard deviations more than unique information, and that groups facing a hidden profile were eight times less likely to reach the correct decision than groups given full information. Mesmer-Magnus and DeChurch [39] confirm across 72 studies that unique information sharing predicts team performance and that teams do it badly by default.

The collective intelligence literature is weaker and we do not lean on it. Woolley and colleagues [55] identify a general factor in small-group performance associated with conversational turn-taking and social sensitivity. Credé and Howardson [13] reanalyse six samples and conclude the general factor explains little variance once low-effort responding and the nested structure of the data are accounted for. A subsequent meta-analysis across 1,356 groups [45] recovers the factor with task loadings between 0.27 and 0.52, predicting held-out task performance at a mean correlation of 0.40. Those loadings are modest, and that defence is authored in part by the original team, so it is not independent replication. We read these results as arguing against a shared-knowledge-pool framing of peer groups and in favour of something closer to a routing and attention-allocation system. The reading is ours: none of the cited studies concerns peer advisory groups, and none proposes a routing account.

2.4 Commitment, and a caution about its evidence

The accountability component of these groups has better empirical support than the rest of it, though less than it had a decade ago. Locke and Latham [36] establish the performance effect of specific and difficult goals. Gollwitzer's implementation intentions [19] show that specifying when, where and how raises follow-through relative to intention alone, and the 2006 meta-analysis of 94 tests reported a medium-to-large effect of d = .65 [20].

That figure should no longer be quoted on its own. The same authors' meta-analysis of 642 tests [48] reports a raw effect of d = .36, identifies publication bias they themselves describe as substantial, and returns a model-averaged d = .15 with a 95% interval from .08 to .22 once Robust Bayesian Meta-Analysis corrects for it. The honest description of implementation intentions today is a small effect, robust in sign, and highly heterogeneous. A design that leans its accountability stage on the 2006 number is leaning on a number its own authors have since revised downward by roughly three quarters.

Progress monitoring has held up better. Harkin and colleagues [23], pooling 138 studies and 19,951 participants, find that interventions prompting people to monitor goal progress raised monitoring frequency (d+ = 1.98) and goal attainment (d+ = 0.40), with monitoring frequency mediating the effect, and with published and unpublished estimates almost identical. They also report larger attainment effects where progress was reported to someone (d+ = 0.47) or made public (d+ = 0.55) than where it was kept private (d+ = 0.19). We rest the argument on the reported-to-someone contrast, which is the better supported of the two, rather than on the public-display estimate. We state that finding in the form the authors state it: this is a comparison across studies rather than a manipulation of publicity, and the authors themselves name experimenter demand among the candidate explanations and call for a direct test. It is suggestive support for a public commitment step, not a demonstration of one.

2.5 What has actually been tested

The design literature above is about mechanisms in general. It is worth stating separately, and bluntly, what is known about this specific intervention, because the answer is uncomfortable for anyone selling it.

We could find no commercial peer advisory group that has been evaluated with a control group. Our search covered Google Scholar, EconLit, PubMed and ProQuest Dissertations for the named providers and for the generic format terms, in English, with no date restriction, to August 2026. Absence of evidence in that search is not proof that no such evaluation exists. We could find no peer-reviewed causal evaluation of Vistage, YPO, EO, or any comparable provider. What exists is a doctoral dissertation interviewing twelve members recruited purposively, by convenience and snowball sampling from the author's own network, the author himself being a member of one such organisation, about their perceptions [17], and vendor-commissioned comparisons of member firms to non-member firms which cannot separate treatment from selection, since firms choose to buy a membership and the traits that drive that choice are the traits that drive growth. This is an evidentiary vacuum, and it is the strongest available argument that a specified alternative deserves testing.

Adjacent to that vacuum sits a small set of genuine field experiments on structured peer contact among firm owners. Table 2 summarises the ones we regard as informative. They are the empirical context this paper is answerable to, and by the reading we give in the final column, five of the six cut against the design proposed here in some respect. Only Ingram and Morris is supportive on balance, and it supports the diagnosis rather than the remedy; even that row records a null on demographic homophily which corrects a story the design might otherwise have leaned on.

Study	Design	Finding	What it does to this proposal
Cai and Szeidl (2018) [8]	1,500 Chinese SME managers randomized into monthly meeting groups of 10 for 12 months; 1,320 controls.	Sales +8.1% at midline and +10.3% at endline, the latter a year after meetings ended. Management practice +0.21 SD. Firms assigned 10 log points larger peers gained an extra 1.05 log points of sales. That figure is	Supports the format at a group size of 10 and a monthly cadence. Cuts against equal composition, since a substantial part of the gain came from being placed with better peers. The gain is not, however, attributable to the funding channel: the authors

Study	Design	Finding	What it does to this proposal
		from the published version; the earlier working paper reports 1.18 at midline and 1.66 at endline for the same quantity.	test it directly and report that "it is not information about the government grant which drives our main results". Their management-practice effect of 0.21 SD points to learning.
Chatterji, Delecourt, Hasan and Koning (2019) [10]	100 Indian high-growth tech founders randomized into 50 pairs at a three-day retreat. There is no untreated control arm: every founder was paired, and the estimate is the marginal effect of a one standard deviation rise in the randomly assigned partner's pre-existing Peer Management Index.	A partner one standard deviation higher on that index is associated with 28% more employees and 10 percentage points higher survival at two years. The authors also report that founders paired with the informal-style managers were more likely to fail and that those who survived grew more slowly.	The active ingredient is the specific partner's practice, not peer contact. Founders with MBAs or accelerator experience showed little responsiveness, and they are the population a premium network sells to. This is also the sharpest warning against the present design: assigned pairing is a two-sided lottery, and the same randomisation that produced the upside produced faster failure on the other tail. A schedule that assigns all twenty-eight pairings maximises exposure to both tails, not only the favourable one.
Fafchamps and Quinn (2018) [16]	239 manufacturing managers in Ethiopia, Tanzania and Zambia randomly assigned to business-plan judging committees, most of five or six judges, fixed in membership and meeting three times, creating exogenous ties; 106 further judges served as a non-committee comparison.	Real new ties formed, but diffusion was narrow. Having a committee peer already VAT-registered at baseline raised a manager's own probability of registering by 6.6 percentage points, and the corresponding figure for holding a bank account is 4.4. The authors' footnote 21 calls these the intention-to-treat effects, and approximates the corresponding local average treatment effects at about 12 and 8 percentage points. No significant evidence of diffusion was found in any other management measure. All three countries were reforming VAT legislation during the experiment, a confound the authors flag themselves.	Damaging for any claim that structure alone diffuses practice. A fixed committee meeting three times moved two administrative practices and nothing else, which suggests that sustained intensity rather than placement produced the larger effects elsewhere in this table.
Ingram and Morris (2007) [30]	92 analysed guests at a business mixer wearing electronic badges that logged dyadic encounters, an encounter requiring sustained proximity of at least a minute.	Only 5% said their goal was to cement existing ties, yet pairs with a strong prior relationship were 223% more likely to meet. No significant demographic homophily in the average encounter, though the authors report it operating for some guests at some points in the evening.	The best evidence that open networking fails its own stated purpose. It also corrects the usual story: guests did meet people unlike themselves, so the failure is in forming new ties, not in avoiding difference.
Brooks, Donovan and Johnson (2018) [5]	Inexperienced young female microenterprise owners in Dandora, Nairobi, randomized to mentorship by a local female owner aged over 40 with at least five years in the same	Mentorship raised profits about 20%. The class produced no detectable profit effect (a small, statistically insignificant estimate) despite changing business practices. The mentorship effect	Supports relational over curricular formats, and and is consistent with, though it does not establish, the benefit being contingent on relationships surviving. We found no study demonstrating a design that

Study	Design	Finding	What it does to this proposal
	business and high sector-adjusted profit, a formal business class, or control (roughly 124, 129 and 119).	faded, and the authors report that fade alongside the dissolution of matches without claiming the dissolution caused it.	keeps them alive. We flag a tension we cannot resolve: the treatment here is an explicitly asymmetric pairing of novice with expert, and the mentor asymmetry is the active ingredient, whereas Section 7.2 requires that neither party be designated mentor or recipient. The strongest relational evidence in this table therefore supports a structure this paper rules out.
Dimitriadis and Koning (2022) [14]	Togolese entrepreneurs randomized to a two-hour social-skills module inside a two-day business programme, with the control arm covering the same curriculum at a slower pace over the same roughly twenty contact hours.	Participants matched with 50% more peers, more of them skill-complementary (0.5 predicted complementary ties against 0.2 in control), and monthly profits rose about 20%. A 2,000-run simulation holding each entrepreneur's tie count fixed while randomising who the ties were with drove the simulated difference to near zero, leading the authors to conclude that forming more ties at random does not result in more useful connections.	A competing intervention rather than a complementary one, and the most direct published evidence against the mechanism proposed here. The binding constraint may be individual capacity to use peer ties, which assigning eight people to each other does not supply. Worse for this paper, the authors' simulation is close to a test of assignment itself: holding tie count fixed and randomising targets produced no gain in complementarity. A distance-graded schedule assigns targets from a seating whose residual, by Section 5.3, is of the same order as the quantity it encodes. We have no answer to this beyond the observation that their randomisation is uniform whereas ours is ordered by an estimated axis, and that difference is exactly what remains to be tested.

Table 2. Field experiments on structured peer contact among firm owners. Effect sizes are as reported by the authors, with the two reinterpretations noted in the rows themselves. Every study in this table was checked against a primary source or, where the published text was closed, against the authors' working paper; the register records which.

Three limitations of this evidence base bear directly on the present proposal, and we state them before making any design claim.

The population is wrong. Every causal study above is set in a developing or emerging economy, among micro, small or early-stage firms with low baseline management quality. That is the left tail, where information is scarcest and the returns to any information are largest. There is no randomized evidence that structured peer groups help established owners in a high-income economy, which is precisely the population that buys this format. Chatterji and colleagues supply the warning shot: their founders with formal management training showed no response.

Rotation has never been tested. We found no study that randomizes a meeting schedule, a rotation, or any rule for which members meet which others and when. The nearest evidence points the other way: Cai and Szeidl compared stable monthly groups against one-time cross-group meetings and found the stable groups produced more business partnerships. That is weak evidence against rotation, and the design proposed here must own it. Our answer is that their comparison confounds rotation with intensity, since the cross-group meetings happened once, whereas the schedule in Section 6 keeps every pair inside a group that continues to meet. That answer is a hypothesis, not a rebuttal.

Group size has never been randomized. No experiment we found varies the number of members as a treatment. There is therefore no empirical basis for eight, nor for ten, nor for twelve. The argument for eight in Section 6.3 is a fact about a schedule, and it should not be mistaken for evidence about a group.

2.6 A note on group size

Dunbar's work relating neocortical volume to group size [15] is routinely cited to justify small-group design. We avoid it. Lindenfors, Wartel and Lind [34] re-estimate the underlying regression on updated primate datasets with modern phylogenetic methods and obtain human point estimates ranging from about 16 to 109 with 95% intervals spanning roughly 2 to 520, concluding that specifying any one number by this method is not supportable. Whatever justifies eight, it is not a cognitive constant. Section 6.3 gives an argument that does not depend on one.

3 A model of member and edge value

3.1 Nodes: capability and reach

Let the group be N = {1, ..., n} with n = 8. Each member i arrives with an endowment *P*i. Rather than decomposing the endowment into an unmeasurable list of attributes, we distinguish only the two parts that behave differently under exchange.

$$P_i = \left(K_i , R_i \right) \tag{1}$$

Capability *K*i is what i can do: skills, judgement, accumulated pattern recognition. We take it to transfer slowly, through practice, and to resist verbal transmission; this is an assumption of the model rather than a result, and P4 in Section 9 states the version of it that could be refuted. **Reach** *R*i is what i can access: people, channels, resources, standing. It transfers in a single sentence, at near-zero cost to the giver.

This distinction is deliberately blunt and it does real work. It predicts that the observable early yield of a peer network is almost entirely reach, and that capability gains, where they occur, appear later and are harder to attribute. If that asymmetry holds, a programme promising capability transformation on a six-month horizon is promising the slow quantity on the fast quantity's timetable. We return to this in Section 10.

3.2 Edges

The group's value is not the sum of the endowments, which would describe eight people who happen to share a room. It is

$$M = \Sigma_{i \in N} P_i + \Sigma_{i<j} C_{ij} \tag{2}$$

where Cij is the value generated by the relationship between i and j. For n = 8 there are 28 such relationships. We propose that Cij is multiplicative in three factors.

$$C_{ij} \propto D_{ij} \times T_{ij} \times L_{ij} \tag{3}$$

Definition 1 (Useful difference). *Dij is the non-redundancy of i and j with respect to the group's declared domain: the extent to which each holds information, framing, or access that the other could not readily derive. This is variety in the sense of Harrison and Klein (see Section 2), not demographic separation.*

Definition 2 (Trust). *Tij is the willingness of each to disclose an unresolved problem in its unflattering form, and to receive a disconfirming response without withdrawal.*

Definition 3 (Legibility). *Lij is how accurately i can model j's current situation: what j is trying to do, what j is good at, what j is currently stuck on, which decisions are live, and which opportunities would be relevant.*

Legibility is the addition we consider most important. Our impression is that peer-group marketing rarely mentions it, though we have not surveyed that material systematically and offer the observation as an impression only. The framing we most often encounter holds that members pool knowledge, so that an insight discovered by one becomes available to all. We regard this as largely false, on the strength of the argument below rather than of data. Our proposal is that what transfers between peers is less the lesson than the person: having someone in the room who has been through this specific thing and will say so at the moment the decision is live. That mechanism requires legibility, not shared knowledge. A member does not need to have absorbed what another member learned. They need to know enough about that member to recognise, at the right moment, that this is a call worth making before signing.

3.3 Why the form is multiplicative

Equation (3) is multiplicative rather than additive because the failure modes are conjunctive, and this is the model's only substantive empirical commitment.

If difference is high and trust is absent, the product is zero. This is what we would expect of a well-curated room of strangers, and we offer it as illustration rather than observation: interesting on paper, cordial in practice, productive of nothing. If trust is high and difference is absent, the product is again zero: the long-running group of near-identical members who enjoy each other and, on our account, stop surprising each other. We have not measured this case either. If difference and trust are both high but legibility is absent, the product is zero a third way: warm relationships between people who cannot name what the other is working on, which we take to be the condition of many conference friendships, though we have not measured it.

The model therefore predicts that assembling diverse people is insufficient, that assembling trusted people is insufficient, and that meeting regularly is insufficient without the specific content that builds situational knowledge.

> **Remark 1 (On stopping here).** *It is tempting to extend equation (3) with further factors: interaction intensity, timing relative to decisions, reciprocity. We resist this. Each additional multiplicand buys descriptive coverage at the cost of the model's teeth, and a product of five unmeasurable terms predicts only that zero times anything is zero, which is a property of multiplication rather than of groups. In particular, relevance at the moment of decision is a property of a moment, not of a pair, and including it in a product over pairs is a category error. We treat timing as an architectural problem instead (Section 7.1).*

3.4 What would falsify the multiplicative form

A multiplicative model is only worth stating if it can lose. Ratings of D, T and L on each realised dyad, together with a dyad-level utility measure U, permit a direct comparison of

$$\log U = \beta_0 + \beta_1 \log D + \beta_2 \log T + \beta_3 \log L + \varepsilon \tag{4}$$

against the additive alternative $U = a + bD + cT + dL + \varepsilon$, out of sample. The multiplicative form is distinguished by two signatures. First, it predicts that a near-zero value on any one factor cannot be compensated by high values on the other two, so the interaction terms should carry variance that main effects do not. Second, it predicts a specific pattern of residuals: the additive model should systematically over-predict utility on dyads that are extreme on one factor and low on another.

We flag an identification problem honestly. In the design we propose, D, T and L would be rated by the same respondent who rates U, and typically soon after the same conversation, so common-method variance would be expected to inflate all three associations. A serious test needs at least the D ratings to come from a source independent of the utility rating, for example from profile data collected at selection rather than from the partner after the meeting.

4 The difference-trust tension

The design problem follows from the model, under one assumption we state plainly because the rest of the paper rests on it. We assume that across the edges of a real group, D and T are negatively correlated. We know of no direct measurement of that correlation, and we label it the load-bearing empirical assumption of Sections 4 to 8 rather than an established finding. The homophily literature [38] supports the weaker claim that ties form preferentially between similar people, which yields the assumption only under further premises about which ties are counted and how difference is scored.

This is not a defect of the members. We take it, following the homophily literature [38], that trust accumulates fastest where communication is cheapest, and that communication is cheapest between people who share vocabulary, context and reference points, which is to say between people for whom D is low. On that reading homophily is not laziness but a rational local strategy, since people invest where the immediate return is highest. The reading is ours; the cited work establishes the tie-formation pattern and not the mechanism we attribute to it. The difficulty is that the local optimum is a global failure. We conjecture, without evidence, that a group of eight left to itself develops something like six or seven strong short edges and leaves the remaining twenty-two or twenty-one relationships as pleasantries, and that the edges it abandons are disproportionately the high-difference edges that equation (3) identifies as most valuable. This is the mechanism the design exists to address, and it is untested; P5 in Section 9 states the version of it that could be refuted.

Two natural remedies, we expect, fail.

Exhortation. Instructing members to connect broadly specifies an outcome without a mechanism, and the instruction competes against a real cost that the instruction does not reduce.

Forced early exposure to maximum difference. Pairing each member with their complement in the first month maximises D at a point where T is near zero, and by equation (3) produces approximately nothing. Our expectation, again untested, is that such a pairing produces little in a way that is legible to the participants, who then conclude that cross-pairing is a waste of time and return to their neighbours believing they have evidence for it.

The resolution we propose is that difference should be introduced on a schedule that tracks the accumulation of trust: short edges first, diameters last. This converts an exhortation into a timetable. Two things are then required, and the rest of the paper supplies them in order. Section 5 asks when the members can be placed on a circle at all. Section 6 asks whether the timetable exists once they are.

5 Placing members on the wheel

Section 6 will show that once members are seated on an octagon, the schedule is essentially forced. All of the design freedom therefore sits here, in the seating, and the draft version of this argument simply asserted that members can be placed so that angular distance encodes difference. That assertion is not free, and this section states what it costs.

5.1 The condition under which an exact placement exists

Seating objects so that distance in the arrangement tracks dissimilarity is a classical problem. In one dimension it is *seriation*, and the matrices that admit an exact solution are the Robinson matrices [46]: a symmetric zero-diagonal dissimilarity is Robinson if there is an ordering under which entries increase as one moves away from the diagonal along every row. For Robinson matrices the ordering is recoverable in polynomial time by a spectral algorithm [4], and seriation has a mature cross-disciplinary literature [33] with standard implementations [22].

The octagon is the circular case, and the relevant class is narrower.

> **Definition 4 (Circular Robinson).** *A symmetric zero-diagonal dissimilarity matrix on n objects is circular Robinson if there is a cyclic order under which, for every object i, the function k → D(i, i+k mod n) is*

> *unimodal: read cyclically, each row rises from the diagonal to a single maximum and then falls (Armstrong, Guzmán and Sing Long [3], Proposition 3.7).*

Substantively, the condition says that difference is essentially one-dimensional and wraps around: each member has exactly one antipode, and dissimilarity to everyone else decreases monotonically in both directions from it. When it holds strictly, a compatible circular order can be computed in O(n log n) time and verified in $O(n^2)$ [9], improving the earlier optimal $O(n^2)$ algorithm [3]. We note, following Carmona and colleagues, that several non-equivalent notions of circular Robinson dissimilarity exist in the literature; Definition 4 is the one we use throughout.

The honest reading is that this is a demanding condition, and we do not expect a matrix of human judgements about how different two business owners are to satisfy it in general. Our reason is structural rather than empirical, since we have collected no such matrices: judgements of this kind are plausibly noisy, frequently multidimensional (industry, stage, temperament, geography), and need not be metric. Section 5.3 demonstrates the failure on a worked example.

5.2 What to do when it fails, and why n = 8 is unusually lucky

When the exact condition fails, the octagon stops being an encoding and becomes a fitted approximation. The named method for fitting is *circular unidimensional scaling* [28], which places objects around a closed circular continuum so as to minimise a least-squares loss between observed proximities and the distances the placement implies. The associated psychometric structure is developed further as circular strongly-anti-Robinson approximation [29].

In general this is hard. Arranging objects on equally spaced points of a circle to minimise total weighted edge length is the Minimum Circular Arrangement problem, which is NP-hard by reduction from Minimum Linear Arrangement [18]; the best known approximation factor for the weighted directed version is O(log n log log n) [40]. Even in the easier linear case, optimally fitting a Robinson structure in the maximum-error norm is NP-hard [12], with a factor-16 polynomial-time approximation the best known guarantee [11]. For large n the defensible route is a two-dimensional Laplacian embedding sorted by angle [44]; the central caveat there is worth repeating, since a one-dimensional spectral embedding cannot recover a circular order at all.

None of that applies here, and this is the one place where the small group size is an unambiguous gift. There are only 7!/2 = 2520 distinct circular arrangements of eight labelled members up to rotation and reflection. The problem that is NP-hard in general is, at n = 8, solved exactly by enumeration in well under a second. We therefore propose the following procedure, and we regard reporting its output as a requirement rather than an option.

1. Elicit a symmetric dissimilarity matrix D over the eight members on the group's declared domain, from selection materials rather than from post-meeting ratings.
2. Enumerate all 2520 circular arrangements. For each, compute chord distance and fit dissimilarity as an affine function of it by least squares.
3. Select the arrangement minimising the residual sum of squares. This is the exact optimum, not a heuristic.
4. Report the residual: root mean squared residual, maximum absolute residual, and the Kendall tau-b between the ranking of observed dissimilarities and the ranking of chord distances.
5. Run the compatibility check of Definition 4 on the chosen order and state plainly whether the exact circular Robinson condition holds.

Steps 4 and 5 attach a reported number to the paper's weakest assumption. We are careful below about how little that number licenses. A cohort whose fit is poor has been told something before it meets: the axis of difference chosen for it may not describe it well. We are careful not to overstate what the residual licenses. Section 5.3 shows the fit statistic is not monotone in placement quality, so a poor fit is a prompt to inspect the axis rather than proof that the schedule is distorted, and a good fit is not proof that it is sound.

5.3 A worked example, and what the residual actually costs

We generated a dissimilarity matrix for eight fictional members from a circular base, a second latent dimension (three members share an industry sector, which reduces their non-redundancy regardless of temperament), and uniform rating noise. This is intended to be realistic rather than favourable. Running the procedure of Section 5.2 gives the results in Table 3.

Quantity	Value	Reading
Arrangements enumerated	2520	The exact optimum, found by exhaustion.
Fitted model	0.041 + 0.148 × chord	One seat of separation is worth 0.148 of dissimilarity.
Root mean squared residual	0.137	Prediction error of the fitted relation, about 0.93 of one seat step. Not a seating error.
Maximum absolute residual	0.276	Worst-fitted pair, about 1.9 seat steps of prediction error.
Kendall tau-b, dissimilarity vs chord	+0.58	Positive and far from perfect.
Circular Robinson holds	No, on all eight rows	The exact condition fails everywhere.

Table 3. Output of the exact placement procedure on the worked example. The script that produces this table is included with the paper.

Two consequences deserve to be stated rather than buried. First, the root mean squared residual of 0.137 is very nearly the value of one chord step, 0.148. That residual is a prediction error in units of dissimilarity, not a seating error: divided by the fitted chord step it is about 0.93 chord units of error in the fitted relation. It should not be read as members sitting in the wrong seats. On this matrix no member does, since the enumeration recovers the generative circular order exactly. What the residual measures is how poorly chord distance predicts dissimilarity once a second latent dimension is present, and it is nearly as large as the whole step between adjacent seats. Second, the seating produces a round-seven pairing between two members whose observed dissimilarity is 0.43, which is lower than that of one of the pairs the schedule treats as merely near. The complementary round, which the entire architecture exists to protect, would in this case pair two people who are not in fact each other's complements.

A sharper objection applies to the two figures this example is most often reduced to, namely that the exact condition fails everywhere and that the residual is about the size of one seat step. Both are functions of the sector penalty we chose, which is a constant we set by hand rather than a measured quantity, so we report the sweep instead of the single number. Holding everything else fixed and varying that penalty from 0.00 to 0.60, the ratio of the residual to the step fitted at that same penalty, rather than to the generative constant of 0.17, runs 0.21, 0.36, 0.62, 0.93, 1.33, 1.27, and the number of rows violating circular Robinson runs 0, 0, 8, 8, 8, 7. The value quoted in Table 3 is the fourth of these. A reader is entitled to conclude that we could have produced almost any impression of the method by turning that dial, and we agree. What the sweep also shows is which quantity is robust: the enumeration recovers the true latent order at every penalty up to 0.45, and fails only at 0.60. We do not offer a threshold, because the sweep does not identify one: at 0.45 the sector effect already exceeds three fitted steps and the order is still recovered, so all we can say is that recovery survives well beyond the point where the exact condition fails and breaks somewhere between 0.45 and 0.60. The honest summary is therefore not that the octagon fits badly by some fixed amount, but that ordering is recoverable well past the point where the exact condition breaks. One further feature of the sweep deserves stating because it is inconvenient. The fit statistic does not degrade monotonically: it rises to 1.33 and then falls back to 1.27 at the penalty where the ordering finally breaks. The fit therefore improves slightly at the moment the method fails, which means the residual is not a usable warning signal for placement failure. A practitioner cannot infer from a good fit that the seating is sound. The accompanying script reproduces this table.

We do not claim this case is representative: the matrix was constructed from a circular base plus a sector effect plus noise, so it demonstrates a possibility rather than measuring a frequency. What it does establish is that the exact condition and the recovered placement are different things: circular Robinson structure failed on all eight rows while the enumeration still returned the true order. Failure of the condition is therefore not by itself a reason to distrust a seating, and it bounds what the rest of the paper can claim. The schedule of Section 6 is exact. The placement it consumes is an approximation whose error is of the same order as the quantity being graded, and any cohort running this design should publish that error alongside its results.

6 The rotation schedule

Assume for this section that a placement has been fixed, so that each member sits at a vertex of a regular octagon and chord distance encodes difference. Label the vertices 0 through 7 and define

$$d(i,j) = \min(\,|i-j|\,,\ 8-|i-j|\,) \in \{1,2,3,4\} \tag{5}$$

Adjacency means similarity and d = 4 means complementarity: vertex i and vertex i+4 are the two members least able to derive each other's perspective. The colour wheel is a natural visual encoding, since it has no first hue and no privileged hue, but does have well-defined complements. Members retain their colours; the design goal is explicitly not convergence. The complete graph on these vertices has 28 edges, distributed by distance as 8 at d = 1, 8 at d = 2, 8 at d = 3, and 4 at d = 4.

6.1 A distance-graded one-factorization

A *round* is a pairing of all eight members into four simultaneous pairs, that is, a perfect matching. A partition of all 28 edges into such rounds is a one-factorization. Call a one-factorization **distance-graded** if every round is homogeneous in chord distance, so that all four pairs meeting in a given round sit at the same remove from one another.

The distinction matters because the standard construction does not have this property. The circle method used for round-robin tournament scheduling produces rounds of mixed chord length, which is precisely what the argument of Section 4 says to avoid. A schedule can cover all 28 relationships and still put the hardest pairing in the first month.

> **Proposition 1 (Distance-graded one-factorization of K8).** *The complete graph on eight vertices, embedded on the octagon, admits a one-factorization into seven rounds, each homogeneous in chord distance: two rounds at d = 1, two at d = 2, two at d = 3, and one at d = 4.*

Proof. Partition the edges by distance into the circulant graphs C8(d) for d in {1,2,3,4}. C8(1) is the octagon 0-1-2-3-4-5-6-7-0, an even cycle, which decomposes into the two alternating perfect matchings {01, 23, 45, 67} and {12, 34, 56, 70}. C8(2) is the disjoint union of the four-cycles 0-2-4-6-0 and 1-3-5-7-1; taking alternating edges from each and combining gives {02, 46, 13, 57} and {24, 60, 35, 71}. C8(3) is the single eight-cycle 0-3-6-1-4-7-2-5-0, an even cycle, which decomposes into {03, 61, 47, 25} and {36, 14, 72, 50}. C8(4) consists of the four diameters {04, 15, 26, 37}, which are pairwise disjoint and already form a perfect matching. These seven matchings are edge-disjoint and their union is all 8 + 8 + 8 + 4 = 28 edges. ▪

Table 1 states the resulting schedule. Members traverse all 28 relationships in seven rounds, in non-decreasing order of chord length, ending with the four diametrically opposite pairs. The order is non-decreasing in measured difference only if the placement of Section 5 is exact. On the worked example of Section 5.3 it is not: the most different pair at chord one scores 0.29 against 0.06 for the least different pair at chord two, and the same inversion occurs at both remaining boundaries. The schedule grades seats reliably and difference only as well as the seating fits. The order is non-decreasing rather than strictly increasing, since there are two rounds at each of the first three distances.

Round	Distance	Pairs	What the round is for
1	d = 1	0-1 2-3 4-5 6-7	Adjacent. Easy exchange, tactical depth with someone who already shares the frame.
2	d = 1	1-2 3-4 5-6 7-0	The other half of the near ties. Trust is built here or nowhere.
3	d = 2	0-2 4-6 1-3 5-7	Near. Shared enough frame to be quick, different enough to surprise.
4	d = 2	2-4 6-0 3-5 7-1	Near. First real translation work, still on friendly ground.
5	d = 3	0-3 6-1 4-7 2-5	Far. Each has to explain themselves properly.
6	d = 3	3-6 1-4 7-2 5-0	Far. Assumptions neither knew they held begin to surface.
7	d = 4	0-4 1-5 2-6 3-7	Complements. The maximum-reframing round the schedule exists to protect.

Table 1. The distance-graded rotation. Every member meets every other exactly once across seven rounds, no member is ever unpaired, and no round mixes chord lengths.

6.2 The schedule is essentially forced

A natural objection is that Table 1 is one arbitrary choice among many, and that the grading is therefore decorative. It is not. The construction leaves almost no freedom.

> **Proposition 2 (Uniqueness).** *The complete graph on eight vertices admits exactly two distance-graded one-factorizations. Fixing the round order to be non-decreasing in distance leaves exactly sixteen admissible schedules.*

Proof. Since every round is homogeneous in distance and the edges at a given distance lie in no other distance class, the rounds at distance d must partition C8(d); the count therefore factorises over the distance classes, and we treat each in turn. An even cycle of length 2m has exactly two perfect matchings, and they partition its edges. C8(1) and C8(3) are single eight-cycles, so each contributes exactly one partition into two rounds. C8(2) is two disjoint four-cycles A and B, each with two perfect matchings; a perfect matching of C8(2) selects one from each, giving four, and each such choice determines its complement, so C8(2) contributes two distinct partitions. C8(4) is forced. The count is therefore $1 \times 2 \times 1 \times 1 = 2$. Ordering multiplies by the two ways of ordering the pair of rounds at each of d = 1, 2, 3, giving $2 \times 8 = 16$. ▪

The practical reading is that once the placement is chosen, the schedule follows. All the design freedom in this system lives in Section 5, which is where it should be resisted most carefully.

6.3 Which group sizes admit the schedule

Eight is often defended on aesthetic or logistical grounds. A sharper argument is available.

> **Proposition 3 (Existence).** *The complete graph on 2n vertices, embedded on the cycle, admits a distance-graded one-factorization if and only if 2n is a power of two.*

Proof. Partition the edges by distance into circulants C2n(d) for d = 1, ..., n. The graph C2n(n) is the perfect matching of the n diameters. For d < n, C2n(d) is 2-regular and consists of gcd(2n, d) disjoint cycles, each of length 2n/gcd(2n, d). A 2-regular graph decomposes into two perfect matchings if and only if all of its cycles are even. Necessity holds because a perfect matching of the whole graph must restrict to a perfect matching of each connected component, and a cycle of odd length admits no perfect matching on its own vertices. Sufficiency holds because the edges of an even cycle alternate into two classes, each meeting every vertex of that cycle exactly once, and taking the union of one class from each cycle gives a perfect matching of the whole graph. Hence a distance-graded one-factorization exists if and only if 2n/gcd(2n, d) is even for every d < n. If $2n = 2^k$, then for d < n the quantity gcd(2n, d) is a power of two dividing d, hence at most 2^{k-2}, so the quotient is at least 4 and in particular even. Conversely, if 2n is not a

power of two, write $2n = 2^a m$ with $m > 1$ odd and take $d = 2^a$. Since $m \geq 3$ we have $2^a < 2^{a-1}m = n$, so d is a valid distance class. Then $\gcd(2n, d) = 2^a$ and the cycle length is m, which is odd, so no decomposition exists. ▪

For the small even sizes a designer is most likely to consider, the proposition is restrictive. Six fails, because the distance-two edges form two triangles. Ten fails, because the distance-two edges form two pentagons. Twelve fails at $d = 4$, where the edges form four triangles. The next size that works is sixteen, at which point the number of relationships rises to 120. We have no principled ground for excluding sixteen. Saying that legibility cannot be maintained across 120 relationships would be an appeal to a cognitive limit of exactly the kind Section 2.6 rejects as unsupportable, and we decline to make it. What the proposition establishes is that eight and sixteen both admit the schedule and that six, ten and twelve do not. The choice between eight and sixteen is not settled here.

Eight is therefore the smallest group size above four that admits a fully distance-graded rotation, and the argument for preferring it to sixteen is practical and untested rather than derived. We note carefully what this does and does not establish. It is a fact about the schedule. It is not a proof that eight-person groups outperform seven- or nine-person groups, and we do not claim that. A group of seven or nine can still meet in a well-designed order; what it cannot do is meet in an order that is homogeneous in difference at every round.

> **Remark 2 (On the target).** *Twenty-eight live relationships is a demanding target and we suspect most groups sustain roughly half. It may be more honest to design so that no edge is dead than to design for all twenty-eight to be strong. The schedule guarantees that every edge is instantiated at least once, which is a weaker and more achievable claim than that every edge becomes load-bearing.*

6.4 Relation to prior work

We state carefully what is and is not new here, because the surrounding combinatorics is classical and it would be easy to overclaim.

That the complete graph on an even number of vertices is one-factorable at all is due to Kirkman [31]; the rotating-polygon construction usually met as the circle method is the even-order form of the Walecki decomposition, whose history and correct attribution are set out by Alspach [1]. The standard monograph is Wallis [54]. That an even cycle splits into two perfect matchings, and that the circulant generated by a single jump length consists of $\gcd(2n, d)$ cycles of length $2n/\gcd(2n, d)$, are textbook facts; in the second case the cycles are the cosets of the subgroup generated by d in the integers modulo 2n.

Proposition 3 is not new. It is a corollary of the Stern-Lenz Lemma [50], which characterises the one-factorable circulants as those for which $2n/\gcd(di, 2n)$ is even for at least one generator di, and which is surveyed by Lindner and Street [35]. Since the edges at a given chord distance occur in no other distance class, each class must be one-factored on its own, so the Stern-Lenz condition must hold for every distance separately, and applying it to each singleton class yields the stated equivalence directly. We record the result explicitly because we have not found this formulation in the literature and because it is the precise obstruction governing the construction, and we claim no novelty for the underlying lemma. We missed the lemma ourselves at first, since Stern and Lenz proved it as a tool inside a paper on Steiner triple systems, where a graph-theoretic keyword search does not surface it. It is certainly not obscure to the design-theory community, which devoted a survey to it; we cite that survey precisely so the reader can see the result is well known there.

> **Remark 3 (A near-miss worth naming).** *The condition in Proposition 3 is the complement of a well-known one. Hartman and Rosa [25] showed that K2n admits a one-factorization on which the cyclic group of order 2n acts sharply transitively on vertices if and only if n is not a power of two greater than 2. The surface resemblance is misleading and the two notions are distinct. We confirmed by exhaustive computation that none of the 6,240 one-factorizations of K8 is invariant under rotation by one vertex, consistent with Hartman and Rosa, and that the distance-graded factorization of Proposition 1 is likewise not rotation-invariant. K8 therefore has no cyclic one-factorization while possessing a distance-graded one. We record this because it settles, by exhaustion rather*

> *than by assertion, that the distance-graded condition and the classical cyclic condition are independent rather than variants of one another.*

The applied literature on constraining which pairs meet when is the round-robin scheduling literature, surveyed by Rasmussen and Trick [42]; the canonical instance of such a constraint is Russell's work on balancing carry-over effects [47], the bias arising when a competitor meets an opponent immediately after that opponent has met someone else. A distance-graded schedule fixes the pairing-to-round map rigidly, so it is fair to ask what it does to carry-over. We computed this rather than assuming it. Under the cyclic convention the schedule of Table 1 is **not** carry-over balanced, with a carry-over effect value of 88 against a balanced ideal of 56. It is better on that metric than the circle method applied to the same eight members, but the margin is narrow and must be stated carefully. Comparing 88 against the circle method's as-published 196 pits a schedule whose round order is pinned to non-decreasing distance against one arbitrary ordering of a schedule whose round order is free. Optimising both, the circle method reaches 96 while the distance-graded schedule cannot go below 88 without breaking the grading that defines it. The fair comparison is therefore 88 against 96. We report this as a computed property of the construction, not as a design goal it was built to satisfy, and not as a selling point.

What we did not find anywhere, after searching the design-theory and scheduling literatures, is any treatment of ordering the rounds of a round robin by a similarity measure defined on the pairs. If this paper has a combinatorial contribution, that framing is it, and it is a modest one.

The mathematical claims in Sections 5 and 6 were verified computationally. The schedule of Table 1 was checked to be a valid one-factorization with every round distance-homogeneous; the uniqueness count of Proposition 2 was obtained by enumeration; Proposition 3 was confirmed for all even sizes from 4 to 64 by constructing every circulant and searching directly for a decomposition into two perfect matchings by backtracking, so that the existence claim is tested rather than the parity criterion restated. The sweep does not consult cycle structure at all; cycle walking is used separately, to display the odd cycles that block the three named failure cases at $2n = 6$, 10 and 12. Scripts accompany the paper.

7 Architecture

The model implies four stages, in order. The order is not stylistic: each stage supplies a factor that the next stage multiplies by.

Stage 1: Mapping. Establish the axis of difference and place members on it by the exact procedure of Section 5.2. This determines the geometry and therefore the entire schedule, and Section 6.2 shows that essentially no freedom remains afterwards. On our account, getting this wrong degrades everything downstream, since a wheel whose chord distance does not track difference produces a schedule closer to arbitrary. Section 5.3 qualifies this: on the worked example the ordering was still recovered where the fit was poor, so a bad fit does not by itself establish that the schedule is arbitrary. The residual must be reported, not assumed away.

Stage 2: Trust. Build trust on the short edges first. We expect this stage to feel least like progress, and members who joined for reach to be restless during it. We make no claim about how to compress it. The field evidence bears only on repetition: the design that produced substantial effects [8] ran monthly for a year, and the design that produced almost nothing [16] was a fixed committee meeting three times. That supports repeated contact over one-off contact, and says nothing about whether intensity can substitute for duration, which is untested.

Stage 3: Rotation. Execute the seven rounds of Table 1. The point we regard as crucial, and offer as a design argument rather than a finding, is that rotation puts two people in a room and does not give them anything to do. We expect that complementary people do not readily find common ground, that being close to what makes them complementary, and that left with an open brief they will be pleasant and produce little, after which the group may conclude the mechanism is worthless. We have not observed this and no study we found tests it.

We therefore propose that the far rounds carry a specific task with a deliberate asymmetry of content inside a symmetric structure: each member's job is to restate their partner's problem in their own domain's language, so that what each produces differs while the obligation on each is the same. We expect the distance-four partner explaining a member's bottleneck in unfamiliar terms to do something the member's near neighbours are less able to do, precisely because the near neighbours share the member's vocabulary. We have not tested this and state it as the design's motivating expectation. The intention is that under this brief difference stops being an obstacle to be overcome and becomes the instrument. The brief is symmetric by construction, since both members restate and both are restated, which is the property Section 7.2 argues for.

Stage 4: Accountability. A commitment loop with a hard edge: commit, act, account, learn, recommit. Not an open enquiry into what everyone has been doing, but a specific check: last session you said you would do this by today, did you, yes or no, and if not, why. The group earns the right to discuss anything else only after this.

Two cautions attach to Stage 4. Following the implementation-intention literature, commitments should specify when and where, not only what, but Section 2.4 records that the effect of that specification is now estimated at roughly a quarter of its 2006 value once publication bias is corrected. The stronger warrant is the progress-monitoring evidence, and even there the public-reporting advantage is a between-study moderator rather than a manipulated variable. We continue to regard this stage as the one most likely to matter, and we now regard the size of its effect as genuinely uncertain.

7.1 The availability problem

Everything above is scheduled. We expect the highest-value event in the system not to be.

The value described in Section 3, a member reaching another before an irreversible decision, occurs on the decision's timetable rather than the group's. We assume that advice six months early is largely noise while the same advice shortly before a commitment can be decisive, and that no meeting cadence reliably produces the second, because the meeting is rarely in the right week. This is an assumption about timing that we do not test and have no citation for. This implies a second channel: a persistent, low-ceremony space where live decisions are posted as they arise, with a norm that posting one is ordinary rather than an imposition. On this account the scheduled sessions build legibility while the open channel is what would convert accumulated legibility into value at the moment it is worth something, and a group with only the first would have well-informed members who arrive late. Neither half has been observed; both follow from the timing assumption stated above.

7.2 Symmetry

An interaction should alter both members. The failure we design against, and which we have not measured, is that one member presents and the others advise, converting the network into a star and removing the reciprocity the architecture depends on. Formally, for a pairing (i, j) we want both endowments to increase, with neither party designated mentor and neither designated recipient. This requirement is in tension with one of the clearer findings in the field evidence [8], where the largest gains accrued to firms placed with larger and better peers. A composition rule that maximises symmetry may forfeit the mechanism that produced the biggest measured effects, and we do not know how to have both. We name the trade-off rather than resolve it.

8 Measurement

We argue that satisfaction is the wrong instrument. We expect that groups which have converged report high satisfaction, because agreement is pleasant, and that groups in the productive middle of a difficult rotation may report low satisfaction while working correctly. Both are conjectures offered as motivation, not findings; we know of no study measuring satisfaction against convergence in this setting. We propose instead a behavioural diagnostic.

> **Routing.** Does value move through the network when nobody has asked it to?

The observable form is a class of unprompted sentences. I saw this and thought of you. Speak to her before you decide. Do not hire them yet, he went through this last year. These require the mover to hold an accurate model of another member's situation, to recognise a match without being prompted, and to bear a small cost voluntarily. We propose them as a behavioural indicator of legibility rather than a validated instrument. Whether a group that is merely enjoying itself can produce them is exactly what a trial would have to establish.

Three quantities are countable over a cohort: introductions made and accepted, commitments kept, and opportunities routed. We suggest a concrete threshold, with the warning that its value is proposed rather than calibrated against data. A cohort averaging fewer than one unprompted routing event per member per quarter over its first six months is, on this proposal, a discussion group however enjoyable its sessions. We put the number forward so that it can be argued with and revised, not because any study supports that particular value. Dyad-level ratings of difference, trust, legibility and utility should be collected after each pairing to support the test in Section 3.4, with the difference ratings drawn from selection materials rather than from the partner after the meeting, so that the comparison is not contaminated by common-method variance.

> **Remark 4 (Goodhart).** *These measures are diagnostics, not scores. We expect that if members are ranked on them, low-value introductions will be manufactured and commitments set small enough to be certain of keeping. That expectation applies the familiar argument that a measure adopted as a target ceases to measure well [51]. We have not tested it in this setting. Whether they are read by the facilitator or displayed to the members is therefore a consequential design decision, and we recommend the former.*

9 Falsifiable predictions

Each prediction below is stated with the observation that would refute it. Where existing evidence already bears on one, we say so, including where it bears against.

P1. Multiplicative edge value. Out of sample, the log-linear specification of equation (4) predicts dyadic utility better than the additive alternative, and a near-zero value on any one factor is not compensated by high values on the other two. Refuted if the additive model predicts as well or better, or if the interaction terms carry no variance beyond main effects.

P2. Trust-conditioned complementarity. For high-difference dyads, rated utility increases with trust measured before the interaction, and the interaction term is positive after controlling for main effects. Refuted if high-difference dyads yield the same utility regardless of prior trust.

P3. Sequencing advantage. Cohorts running the distance-graded order report higher utility on the distance-four round than cohorts running a randomised round order. Refuted if the complementary round performs no better under grading. This is the paper's central operational claim and it is entirely untested; we found no study that randomises a meeting schedule.

P4. Reach precedes capability. Observable reach events, meaning introductions, referrals and access, increase earlier in a cohort than self- or peer-rated capability. Refuted if capability appears to move faster than reach, which would falsify the transfer asymmetry in Section 3.1.

P5. Rotation counteracts fragmentation. Given free choice alongside required rotations, free-choice contact concentrates on lower-difference dyads, and the designed schedule raises edge coverage relative to a group with no schedule. Refuted if coverage is the same either way. Note that the nearest existing evidence points the other way at the group level, since stable groups outproduced one-time cross-group meetings in the only comparison we could find.

P6. Routing rises with legibility. The share of useful events initiated without a direct request rises over the life of a cohort. Refuted if unprompted routing stays flat while satisfaction rises, which would indicate that legibility never developed.

10 Limitations

This paper offers no empirical validation. It contains no cohort, no data and no measured outcomes. The mathematics of Sections 5 and 6 is proved and computationally checked; nothing else in it should be read as an established result. Eight specific weaknesses deserve naming, each with the observation that would expose it.

The population is wrong. Every randomized study bearing on this design was run among micro, small or early-stage firms in developing or emerging economies, where baseline management quality is low and the returns to any information are large. The format is sold to established owners in high-income economies. Exposed by: a trial in that population showing null effects, which is what Chatterji and colleagues already observed for the subgroup of founders with formal management training.

The axis is assumed measurable, and the fit is poor. Stage 1 requires placing members on a single circular axis such that angular distance tracks difference. Real difference is high-dimensional, and Section 5.3 shows a simulated case in which the exact condition fails on every row while the fitted relation carries a prediction error close to one seat step. We report there how sensitive both of those figures are to a constant we set by hand. Exposed by: two independent facilitators producing materially different seatings from the same selection materials, which would show the placement is not a measurement at all.

The factors are unmeasured. Difference, trust and legibility are defined but not operationalised, and the multiplicative form is motivated by failure modes rather than fitted to data. Section 3.4 also concedes a common-method problem that a naive test would not escape. Exposed by: a properly instrumented study in which the additive model predicts as well.

Selection may dominate. The architecture assumes a group can be composed with genuine difference, sufficient common purpose and no status gradient. It is entirely possible that composition accounts for most of the variance and that the schedule is a second-order effect. The field evidence leans this way, since the largest measured gains came from who members were placed with rather than from how the meetings ran. Exposed by: a design that randomises schedule while holding composition fixed and finds nothing.

The schedule is untested and the nearest evidence is unfavourable. No study randomises rotation. The one comparison we found favoured stable groups over rotating contact, and our reading that it confounds rotation with intensity is a hypothesis rather than a rebuttal. Exposed by: a sequencing trial in which randomised round order performs as well as graded order.

The accountability warrant has weakened. The stage we regard as most likely to matter rests partly on an effect its own authors have revised from d = .65 to a bias-corrected d = .15, and partly on a moderator comparison that was never manipulated. Exposed by: a direct test of public versus private commitment reporting inside a peer group finding no difference.

The schedule grades chord, not measured difference. Section 6 orders rounds by chord length, and chord length equals difference only when the placement of Section 5 is exact. We have no reason to expect exactness on real data, and it failed on the one simulated matrix we built, but we have measured no real seating and so cannot say how often or how badly it fails. On the worked example of Section 5.3 the ordering inverts at every boundary: the most different pair at chord one exceeds the least different pair at chord two, and the same holds at the two remaining boundaries. The schedule is therefore exactly graded in seats and only approximately graded in the quantity the design cares about, and the size of that gap is a property of the data rather than of the schedule. Exposed by: a cohort in which the round-by-round ordering of measured difference shows no monotone trend at all, which would mean the seating carries no signal and the grading is decorative.

The transformation chain is not claimed. A tempting story holds that an insight from one member travels through several hands and returns improved. We regard this as narrative rather than mechanism. Ideas mutate through transmission and may as easily degrade as improve, a pattern the serial-reproduction literature has studied since Bartlett though we do not cite a specific estimate here, and we are aware of no method that would distinguish the

two from inside the group. Related to this, the model predicts that reach transfers quickly and is attributable within months while capability and decision quality are not, which is P4 in Section 9 and is untested. If that prediction holds, any claim about improved decision quality on a six-month horizon is unfalsifiable within that horizon and should not be made. Exposed by: a cohort in which members report the transformation-chain experience at high rates while traced insight histories show no such chain, which would demonstrate that the architecture is generating a satisfying narrative rather than the mechanism, and that participant testimony about this design cannot be trusted as evidence for it.

11 Conclusion

The design principle compresses to four instructions.

> **Protect the colours.** Members should not converge. Overlap should occur in legibility, not identity. The intended end state, which we have not observed in any cohort, is that the green member comes to understand the magenta member extremely well and remains green.
>
> **Strengthen the edges.** On the model proposed here, value lives in the twenty-eight relationships rather than in the eight endowments.
>
> **Rotate the chords.** Traverse all edges on a schedule that grades difference against accumulated trust, ending at the complements.
>
> **Sharpen the rings.** We suggest, as a design intention rather than a claim about people, that the aim is a better resolved field of influence rather than a larger one: knowing more precisely where you have agency, who you can help, and which decisions deserve attention. Nothing in the model above establishes this and we offer it as the value the architecture is built to serve.

And the definition the architecture is built to serve: a mastermind is a small network of autonomous people, designed so that each member becomes sufficiently known, trusted and connected that the growth, experience and opportunities of one increase the agency of the others.

The design question is correspondingly sharp. What is the minimum architecture required to keep twenty-eight human relationships sufficiently alive that eight independent people genuinely make one another more agentic? This paper proposes an answer, and is careful about which parts of that answer are established. The schedule exists and is essentially unique. The sizes admitting it are characterised, by a corollary of a lemma from 1980. The seating problem is exactly solvable at this size. On the one simulated matrix we built, the fitted relation carries an error comparable to the quantity it encodes, and we have no data on how large that error is in a real group, because we collected none. Everything past that is a prediction. Testing it requires cohorts, instruments and time, and we could find no published study supplying all three, though our search was bounded and Section 2.5 says so.

Appendix A Computational verification

Three stdlib-only Python scripts accompany this paper and reproduce every computed claim in it. They are written so that each check is independent of the argument it verifies. Cycle structure is obtained by walking the graph rather than by applying the gcd formula the proofs use; existence at each size is established by searching for a decomposition into two perfect matchings rather than by restating the parity criterion; and the uniqueness count is confirmed a second time by direct enumeration of distance-graded factorizations, which does not reuse the multiplicative structure of the proof.

Script	What it establishes	Result
verify_math.py	Table 1 is a valid one-factorization with every round distance-homogeneous; the K8 distance census is 8/8/8/4; Proposition 2's uniqueness	All checks pass. Exactly 2 distance-graded one-factorizations of K8, agreeing between the per-class count and a direct enumeration, and 16

Script	What it establishes	Result
	count, obtained twice by independent routes; Proposition 3 by exhaustive matching search for every even size from 4 to 64.	schedules once the round order is fixed non-decreasing. The construction agrees with the power-of-two criterion at all 31 sizes tested.
verify_priorart.py	That K8 has 105 perfect matchings and 6,240 one-factorizations; that none is invariant under rotation by one vertex; the carry-over profile of Table 1 against the circle method, both as published and with each schedule's round order optimised.	All checks pass. No cyclic one-factorization of K8 exists, consistent with Hartman and Rosa. Carry-over effect value 88 for the distance-graded schedule against 196 for the circle method as published, neither balanced. Optimising the round order of both, the fair comparison is 88 against 96, and Section 6.4 rather than this figure is the one to quote.
place.py	The exact placement procedure of Section 5.2 on the worked example: enumeration of all 2520 circular arrangements, least-squares fit, residual reporting, the circular Robinson compatibility check, and the sensitivity sweep of Section 5.3 over the sector penalty, which uses the same unimodality test as the main run.	Optimal seating found by exhaustion. Root mean squared residual 0.137 against a fitted chord step of 0.148. Kendall tau-b +0.58. The exact circular Robinson condition fails on all eight rows.

Table 4. Computational verification of the paper's mathematical and numerical claims.

References

[1] Alspach, B. (2008). The wonderful Walecki construction. Bulletin of the Institute of Combinatorics and its Applications, 52, 7–20.

[2] Aral, S., & Van Alstyne, M. (2011). The Diversity-Bandwidth Trade-off. American Journal of Sociology, 117(1), 90–171. https://doi.org/10.1086/661238

[3] Armstrong, S., Guzmán, C., & Sing Long, C. A. (2021). An optimal algorithm for strict circular seriation. arXiv:2106.05944.

[4] Atkins, J. E., Boman, E. G., & Hendrickson, B. (1998). A Spectral Algorithm for Seriation and the Consecutive Ones Problem. SIAM Journal on Computing, 28(1), 297–310. https://doi.org/10.1137/S0097539795285771

[5] Brooks, W., Donovan, K., & Johnson, T. R. (2018). Mentors or Teachers? Microenterprise Training in Kenya. American Economic Journal: Applied Economics, 10(4), 196–221. https://doi.org/10.1257/app.20170042

[6] Burt, R. S. (1992). Structural Holes: The Social Structure of Competition. Cambridge, MA: Harvard University Press. ISBN 0-674-84372-X

[7] Burt, R. S. (2004). Structural Holes and Good Ideas. American Journal of Sociology, 110(2), 349–399. https://doi.org/10.1086/421787

[8] Cai, J., & Szeidl, A. (2018). Interfirm Relationships and Business Performance. The Quarterly Journal of Economics, 133(3), 1229–1282. https://doi.org/10.1093/qje/qjx049

[9] Carmona, M., Chepoi, V., Naves, G., & Préa, P. (2023). A simple and optimal algorithm for strict circular seriation. SIAM Journal on Mathematics of Data Science, 5(1), 201–221. https://doi.org/10.1137/22M1495342

[10] Chatterji, A., Delecourt, S., Hasan, S., & Koning, R. (2019). When does advice impact startup performance? Strategic Management Journal, 40(3), 331–356. https://doi.org/10.1002/smj.2987

[11] Chepoi, V., & Seston, M. (2011). Seriation in the Presence of Errors: A Factor 16 Approximation Algorithm for l-infinity-Fitting Robinson Structures to Distances. Algorithmica, 59(4), 521–568. https://doi.org/10.1007/s00453-009-9319-y

[12] Chepoi, V., Fichet, B., & Seston, M. (2009). Seriation in the Presence of Errors: NP-hardness of l-infinity Fitting Robinson Structures to Dissimilarity Matrices. Journal of Classification, 26(3), 279–296. https://doi.org/10.1007/s00357-009-9041-0

[13] Credé, M., & Howardson, G. (2017). The structure of group task performance: A second look at collective intelligence. Comment on Woolley et al. (2010). Journal of Applied Psychology, 102(10), 1483–1492. https://doi.org/10.1037/apl0000176

[14] Dimitriadis, S., & Koning, R. (2022). Social Skills Improve Business Performance: Evidence from a Randomized Control Trial with Entrepreneurs in Togo. Management Science, 68(12), 8635–8657. https://doi.org/10.1287/mnsc.2022.4334

[15] Dunbar, R. I. M. (1993). Coevolution of neocortical size, group size and language in humans. Behavioral and Brain Sciences, 16(4), 681–694. https://doi.org/10.1017/S0140525X00032325

[16] Fafchamps, M., & Quinn, S. (2018). Networks and Manufacturing Firms in Africa: Results from a Randomized Field Experiment. The World Bank Economic Review, 32(3), 656–675. https://doi.org/10.1093/wber/lhw057

[17] Feghali, A. (2022). Executive Peer Advisory Groups: Who They Are? What Are Their Benefits? Why Do Members Join and Stay? Doctoral dissertation, University of San Diego.

[18] Gansner, E. R., & Koren, Y. (2006). Improved Circular Layouts. In Graph Drawing 2006, Lecture Notes in Computer Science 4372, 386–398.

[19] Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. American Psychologist, 54(7), 493–503. https://doi.org/10.1037/0003-066X.54.7.493

[20] Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. Advances in Experimental Social Psychology, 38, 69–119. https://doi.org/10.1016/S0065-2601(06)38002-1

[21] Granovetter, M. S. (1973). The Strength of Weak Ties. American Journal of Sociology, 78(6), 1360–1380. https://doi.org/10.1086/225469

[22] Hahsler, M., Hornik, K., & Buchta, C. (2008). Getting Things in Order: An Introduction to the R Package seriation. Journal of Statistical Software, 25(3), 1–34. https://doi.org/10.18637/jss.v025.i03

[23] Harkin, B., Webb, T. L., Chang, B. P. I., Prestwich, A., Conner, M., Kellar, I., Benn, Y., & Sheeran, P. (2016). Does monitoring goal progress promote goal attainment? A meta-analysis of the experimental evidence. Psychological Bulletin, 142(2), 198–229. https://doi.org/10.1037/bul0000025

[24] Harrison, D. A., & Klein, K. J. (2007). What's the Difference? Diversity Constructs as Separation, Variety, or Disparity in Organizations. Academy of Management Review, 32(4), 1199–1228. https://doi.org/10.5465/amr.2007.26586096

[25] Hartman, A., & Rosa, A. (1985). Cyclic one-factorization of the complete graph. European Journal of Combinatorics, 6(1), 45–48. https://doi.org/10.1016/S0195-6698(85)80020-2

[26] Hill, N. (1937). Think and Grow Rich. Meriden, CT: The Ralston Society.

[27] Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. PNAS, 101(46), 16385–16389. https://doi.org/10.1073/pnas.0403723101

[28] Hubert, L., Arabie, P., & Meulman, J. (1997). Linear and circular unidimensional scaling for symmetric proximity matrices. British Journal of Mathematical and Statistical Psychology, 50(2), 253–284. https://doi.org/10.1111/j.2044-8317.1997.tb01145.x

[29] Hubert, L., Arabie, P., & Meulman, J. (1998). Graph-Theoretic Representations for Proximity Matrices Through Strongly-Anti-Robinson or Circular Strongly-Anti-Robinson Matrices. Psychometrika, 63(4), 341–358. https://doi.org/10.1007/BF02294859

[30] Ingram, P., & Morris, M. W. (2007). Do People Mix at Mixers? Structure, Homophily, and the "Life of the Party." Administrative Science Quarterly, 52(4), 558–585. https://doi.org/10.2189/asqu.52.4.558

[31] Kirkman, T. P. (1847). On a problem in combinations. The Cambridge and Dublin Mathematical Journal, 2, 191–204.

[32] Kuehn, D. (2017). Diversity, Ability, and Democracy: A Note on Thompson's Challenge to Hong and Page. Critical Review, 29(1), 72–87. https://doi.org/10.1080/08913811.2017.1288455

[33] Liiv, I. (2010). Seriation and matrix reordering methods: An historical overview. Statistical Analysis and Data Mining, 3(2), 70–91. https://doi.org/10.1002/sam.10071

[34] Lindenfors, P., Wartel, A., & Lind, J. (2021). 'Dunbar's number' deconstructed. Biology Letters, 17(5), 20210158. https://doi.org/10.1098/rsbl.2021.0158

[35] Lindner, C. C., & Street, A. P. (1991). The Stern-Lenz Theorem: background and applications. Congressus Numerantium, 80, 5–21.

[36] Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. American Psychologist, 57(9), 705–717. https://doi.org/10.1037/0003-066X.57.9.705

[37] Lu, L., Yuan, Y. C., & McLeod, P. L. (2012). Twenty-Five Years of Hidden Profiles in Group Decision Making: A Meta-Analysis. Personality and Social Psychology Review, 16(1), 54–75. https://doi.org/10.1177/1088868311417243

[38] McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. Annual Review of Sociology, 27, 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

[39] Mesmer-Magnus, J. R., & DeChurch, L. A. (2009). Information sharing and team performance: A meta-analysis. Journal of Applied Psychology, 94(2), 535–546. https://doi.org/10.1037/a0013773

[40] Naor, J. S., & Schwartz, R. (2010). The directed circular arrangement problem. ACM Transactions on Algorithms, 6(3), 1–22. https://doi.org/10.1145/1798596.1798600

[41] Page, S. E. (2007). The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies. Princeton, NJ: Princeton University Press. ISBN 9780691128382

[42] Rasmussen, R. V., & Trick, M. A. (2008). Round robin scheduling: a survey. European Journal of Operational Research, 188(3), 617–636. https://doi.org/10.1016/j.ejor.2007.05.046

[43] Reagans, R., & McEvily, B. (2003). Network Structure and Knowledge Transfer: The Effects of Cohesion and Range. Administrative Science Quarterly, 48(2), 240–267. https://doi.org/10.2307/3556658

[44] Recanati, A., Kerdreux, T., & d'Aspremont, A. (2018). Reconstructing Latent Orderings by Spectral Clustering. arXiv:1807.07122.

[45] Riedl, C., Kim, Y. J., Gupta, P., Malone, T. W., & Woolley, A. W. (2021). Quantifying collective intelligence in human groups. PNAS, 118(21), e2005737118. https://doi.org/10.1073/pnas.2005737118

[46] Robinson, W. S. (1951). A Method for Chronologically Ordering Archaeological Deposits. American Antiquity, 16(4), 293–301. https://doi.org/10.2307/276978

[47] Russell, K. G. (1980). Balancing carry-over effects in round robin tournaments. Biometrika, 67(1), 127–131. https://doi.org/10.1093/biomet/67.1.127

[48] Sheeran, P., Listrom, O., & Gollwitzer, P. M. (2025). The when and how of planning: Meta-analysis of the scope and components of implementation intentions in 642 tests. European Review of Social Psychology, 36(1), 162–194. https://doi.org/10.1080/10463283.2024.2334563

[49] Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. Journal of Personality and Social Psychology, 48(6), 1467–1478. https://doi.org/10.1037/0022-3514.48.6.1467

[50] Stern, G., & Lenz, H. (1980). Steiner triple systems with given subspaces: another proof of the Doyen-Wilson theorem. Bollettino della Unione Matematica Italiana, Serie A, 17, 109–114.

[51] Strathern, M. (1997). 'Improving ratings': audit in the British University system. European Review, 5(3), 305–321. https://doi.org/10.1002/(SICI)1234-981X(199707)5:3<305::AID-EURO184>3.0.CO;2-4

[52] Thompson, A. (2014). Does Diversity Trump Ability? An Example of the Misuse of Mathematics in the Social Sciences. Notices of the American Mathematical Society, 61(9), 1024–1030. https://doi.org/10.1090/noti1163

[53] Uzzi, B. (1997). Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness. Administrative Science Quarterly, 42(1), 35–67. https://doi.org/10.2307/2393808

[54] Wallis, W. D. (1997). One-Factorizations. Mathematics and Its Applications, vol. 390. Dordrecht: Kluwer Academic Publishers.

[55] Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a Collective Intelligence Factor in the Performance of Human Groups. Science, 330(6004), 686–688. https://doi.org/10.1126/science.1193147

Verification note. Every reference above was checked on 22 August 2026 against either the work itself or, where the work is not openly readable, a named secondary record; the register states which for every entry and records the URL consulted. Forty-four entries were verified against the work itself. Ten were not, and are labelled VERIFIED-SECONDARY in the register: [1] Alspach, [5] Brooks and colleagues, [10] Chatterji and colleagues, [11] Chepoi and Seston, [25] Hartman and Rosa, [26] Hill, [35] Lindner and Street, [41] Page, [50] Stern and Lenz, and [53] Wallis. For those, the bibliographic record and the content attributed to the work were confirmed against zbMATH review records, Crossref metadata, publisher or catalogue records, or an author working paper, each named in the register. Two deserve flagging here rather than only in the register. The Chatterji effect sizes and the failure result attributed to passive partners come from the authors' NBER working paper, the published text being unreachable. And [50] Stern and Lenz, on which the paper's central novelty disclaimer rests, is corroborated only for the statement of the lemma, which appears verbatim in a lecture text, and not for its bibliographic record, which rests on the attribution being standard in the design-theory literature and on the survey at [35]. Kirkman was read directly from a scan of the 1847 volume.