

Pull-out: level design workbook

Umut Eren Demirci · October 2026

Every number here is measured from the 40 shipped levels (Assets/_Project/Resources/Levels.txt) by Tools/report.py, which plays the game's own rules. Nothing is estimated. Blue cells are inputs; the formulas update when you change them.

Sheet	What it shows
Levels	Every level: its shape (mat, pencils, colours, boxes, bands, clips, blots, the clock), its par and star thresholds, the tray it needs, how open it is (opening odds, pencils blocked at the start), what its newest mechanic adds, and how two bots fared.
Chapters	Each chapter's summary as formulas over the Levels sheet, and how the measures move together (correlations).
Tuning	The star rule's inputs and the clock's (seconds to start, seconds a pencil). Change them and every threshold and clock follows.
Findings	What the solver and the bots changed in the design, with the measurement behind each change.

Measures

Measure	Meaning
Par	The fewest pencils parked in the tray over any winning line. Every win takes one tap per pencil, so taps can't score a level; parking is the cost. Found by a search over (pencils pulled, pencils boxed) with one queue per cost, as a banded pull can park two.
Tray needed	The smallest tray that still lets the level be won. 5 means the full tray: one careless park and the level is lost.
Opening odds	The share of the first pulls available that lie on some best line.
Blocked at start	Pencils that can't come out on the first tap: something lies in the way or a clip holds them.
What it adds	The level's newest mechanic switched off: bands cut, clips taken off or ink wiped. Par with it minus par without it, and the tray needed with it minus without it. Every level with a focus changes at least one of the two: the generator keeps no deal where the mechanic is decoration.
Random bot	Pulls any pencil it may. 300 runs a level, capped at 400 pulls. A run ends in a dead end when no pull is possible. Median parked is over the runs that won; the 3-star share is over all runs. Neither bot plays the clock.
Greedy bot	Boxes a pencil when it can, otherwise pulls at random; looks no further. A floor for a careless player, not a model of a person.

Tuning: the star rule and the clock

Blue cells are inputs; the Levels sheet reads them.

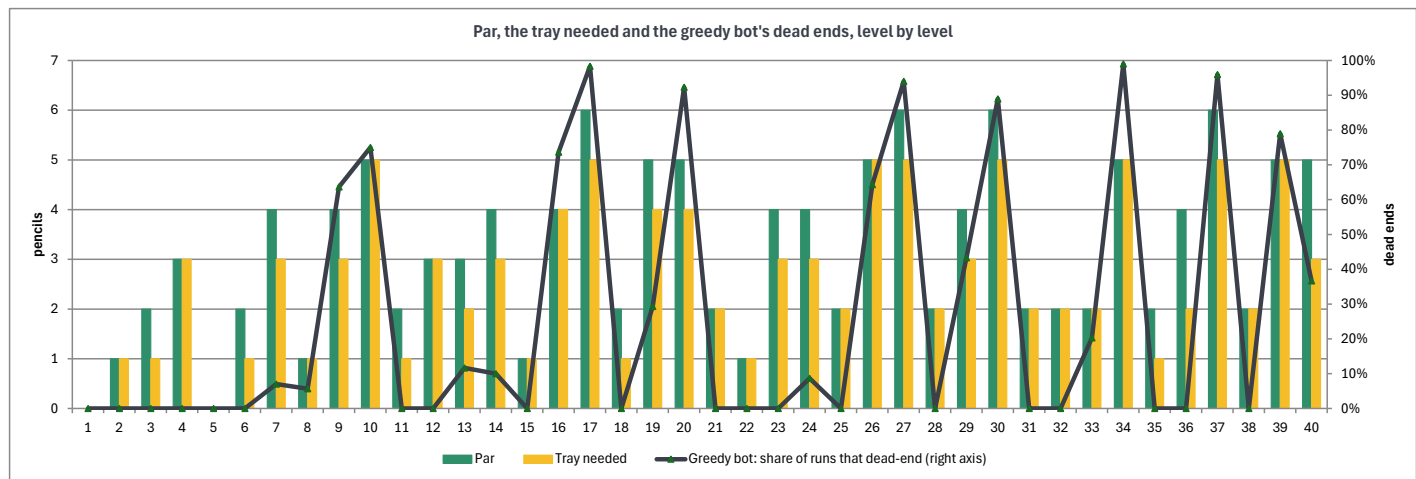
Input	Value	In the game
3 stars: at most par + this many parked		0 Stars.For: parked <= par
2 stars: at most par + this many parked		2 Stars.For: parked <= par + 2
Clock: seconds to start		15 Tools/levels.py time_for
Clock: seconds a pencil		3 Tools/levels.py time_for

The game uses 0 and 2, and 15 s plus 3 s a pencil on levels 10, 15, 20 and so on to 40; the clock starts with the first pull. A win that used undo is capped at two stars, since undo makes par reachable by trial and error. Out of time offers a retry or 15 more seconds once. Whether 3 s a pencil is fair is a playtest question (see Findings).

Every level, measured

From Tools/report.py over the shipped level file. Green = easier, red = harder.

Level	Ch.	Name	Mat	Pencils	Colours	Boxes	Bands	Clips	Blots	Timed	Clock (s)	Focus	Par	3 stars up to	2 stars up to	First pulls	On a best line	Opening odds	Blocked at start	Tray needed	Par without focus	Parked it adds	Tray needed without focus	Tray it adds	Random bot: dead ends	Random bot: median parked	Random bot: 3-star share	Greedy bot: dead ends	Greedy bot: median parked	Greedy bot: 3-star share
1	1	First Pull	4x5	6	2	2	0	0	0				0	0	2	3	2	67%	3	0					0%	1	15%	0%	0	100%
2	1	The Tray	4x5	6	2	2	0	0	0				1	1	3	4	3	75%	2	1					0%	2	13%	0%	2	49%
3	1	Three Colours	5x6	9	3	3	0	0	0				2	2	4	4	2	50%	5	1					0%	4	0%	0%	4	11%
4	1	Full House	5x6	9	3	3	0	0	0				3	3	5	5	4	80%	4	3					0%	4	4%	0%	4	20%
5	1	Breather	4x5	6	2	2	0	0	0				0	0	2	3	2	67%	3	0					0%	1	15%	0%	0	100%
6	1	Six of One	5x6	12	3	4	0	0	0				2	2	4	5	4	80%	7	1					0%	5	1%	0%	3	21%
7	1	Crossings	5x7	12	4	4	0	0	0				4	4	6	6	5	83%	6	3					3%	6	0%	7%	5	13%
8	1	Short Break	5x5	9	3	3	0	0	0				1	1	3	5	3	60%	4	1					7%	5	0%	6%	2	34%
9	1	Long Pencils	6x7	12	4	4	0	0	0				4	4	6	4	3	75%	8	3					63%	6	0%	64%	5	4%
10	1	Pencil Case	6x7	15	4	5	0	0	0	yes	60		5	5	7	4	4	100%	11	5					98%	6	0%	75%	5	23%
11	2	Banded	5x6	9	3	3	1	0	0			bands	2	2	4	5	2	40%	3	1	0	+2	0	+1	0%	5	1%	0%	4	11%
12	2	Two by Two	5x6	9	3	3	2	0	0			bands	3	3	5	4	3	75%	4	3	3	0	2	+1	0%	5	3%	0%	4	17%
13	2	Stretch	5x7	12	4	4	2	0	0			bands	3	3	5	5	5	100%	6	2	2	+1	2	0	4%	5	1%	12%	4	17%
14	2	Snap	5x7	12	3	4	3	0	0			bands	4	4	6	5	3	60%	6	3	1	+3	1	+2	13%	6	5%	10%	5	0%
15	2	Elastic	5x5	9	3	3	1	0	0	yes	42	bands	1	1	3	3	3	100%	6	1	0	+1	0	+1	0%	4	2%	0%	2	48%
16	2	Bundles	6x7	12	4	4	3	0	0			bands	4	4	6	2	2	100%	10	4	2	+2	2	+2	83%	5	0%	74%	5	8%
17	2	Five Colours	6x7	15	5	5	3	0	0			bands	6	6	8	6	3	50%	9	5	4	+2	3	+2	100%	8	0%	98%	6	2%
18	2	Loose Ends	5x6	9	3	3	2	0	0			bands	2	2	4	4	2	50%	5	1	1	+1	1	0	0%	4	1%	0%	3	33%
19	2	Tangle	6x8	15	4	5	4	0	0			bands	5	5	7	5	4	80%	9	4	3	+2	3	+1	32%	7	0%	29%	6	25%
20	2	Rubber Room	6x8	15	5	5	4	0	0	yes	60	bands	5	5	7	4	2	50%	8	4	3	+2	1	+3	89%	8	0%	92%	6	3%
21	3	Clipped	5x6	9	3	3	0	1	0			clips	2	2	4	5	5	100%	4	2	1	+1	1	+1	0%	5	0%	0%	3	36%
22	3	Hold On	5x6	9	3	3	0	2	0			clips	1	1	3	4	4	100%	5	1	0	+1	0	+1	0%	4	3%	0%	1	100%
23	3	Waiting Room	5x7	12	4	4	0	2	0			clips	4	4	6	6	6	100%	6	3	3	+1	3	0	0%	7	1%	0%	4	54%
24	3	Bands and Clips	5x7	12	4	4	1	2	0			clips	4	4	6	4	4	100%	7	3	2	+2	2	+1	1%	6	1%	9%	4	48%
25	3	Quick Clip	5x5	9	3	3	0	1	0	yes	42	clips	2	2	4	5	5	100%	4	2	0	+2	0	+2	0%	5	1%	0%	2	100%
26	3	Paper Trail	6x7	15	4	5	0	3	0			clips	5	5	7	2	2	100%	13	5	4	+1	4	+1	56%	6	6%	64%	5	36%
27	3	Office Hours	6x7	15	5	5	0	3	0			clips	6	6	8	5	5	100%	10	5	5	+1	5	0	97%	8	0%	94%	7	2%
28	3	Coffee Break	5x6	9	3	3	0	2	0			clips	2	2	4	2	2	100%	7	2	0	+2	0	+2	0%	4	2%	0%	2	100%
29	3	Stapled	6x8	15	5	5	2	3	0			clips	4	4	6	3	3	100%	11	3	3	+1	2	+1	37%	7	0%	43%	6	9%
30	3	In Tray	6x8	18	5	6	0	4	0	yes	69	clips	6	6	8	6	6	100%	12	5	3	+3	2	+3	95%	9	0%	89%	8	2%
31	4	Ink Blot	5x6	9	3	3	0	0	1			ink	2	2	4	1	1	100%	8	2	1	+1	1	+1	0%	4	1%	0%	3	32%
32	4	Spill	5x6	9	3	3	0	0	2			ink	2	2	4	3	2	67%	6	2	1	+1	1	+1	0%	5	1%	0%	3	23%
33	4	Smudges	5x7	12	4	4	0	0	2			ink	2	2	4	5	4	80%	7	2	1	+1	1	+1	11%	6	0%	20%	4	17%
34	4	Blotting Paper	6x7	12	4	4	0	0	3			ink	5	5	7	4	3	75%	8	5	4	+1	4	+1	99%	7	0%	99%	5	1%
35	4	Dry Ink	5x5	9	3	3	0	0	1	yes	42	ink	2	2	4	5	4	80%	4	1	1	+1	1	0	0%	4	2%	0%	3	38%
36	4	Ink and Bands	6x7	12	4	4	2	0	2			ink	4	4	6	1	1	100%	10	2	2	+2	2	0	0%	6	2%	0%	4	65%
37	4	Ink and Clips	6x7	15	5	5	0	2	2			ink	6	6	8	3	3	100%	12	5	4	+2	4	+1	97%	6	2%	96%	6	4%
38	4	Recess	5x6	9	3	3	1	0	1			ink	2	2	4	2	2	100%	7	2	1	+1	1	+1	0%	3	25%	0%	2	100%
39	4	Exam Day	6x8	15	5	5	2	2	3			ink	5	5	7	4	2	50%	10	5	3	+2	2	+3	88%	7	0%	79%	6	1%
40	4	Head Teacher	7x8	18	5	6	2	2	3	yes	69	ink	5	5	7	4	3	75%	13	3	4	+1	3	0	31%	8	0%	37%	6	5%



Chapters: the shape of the curve

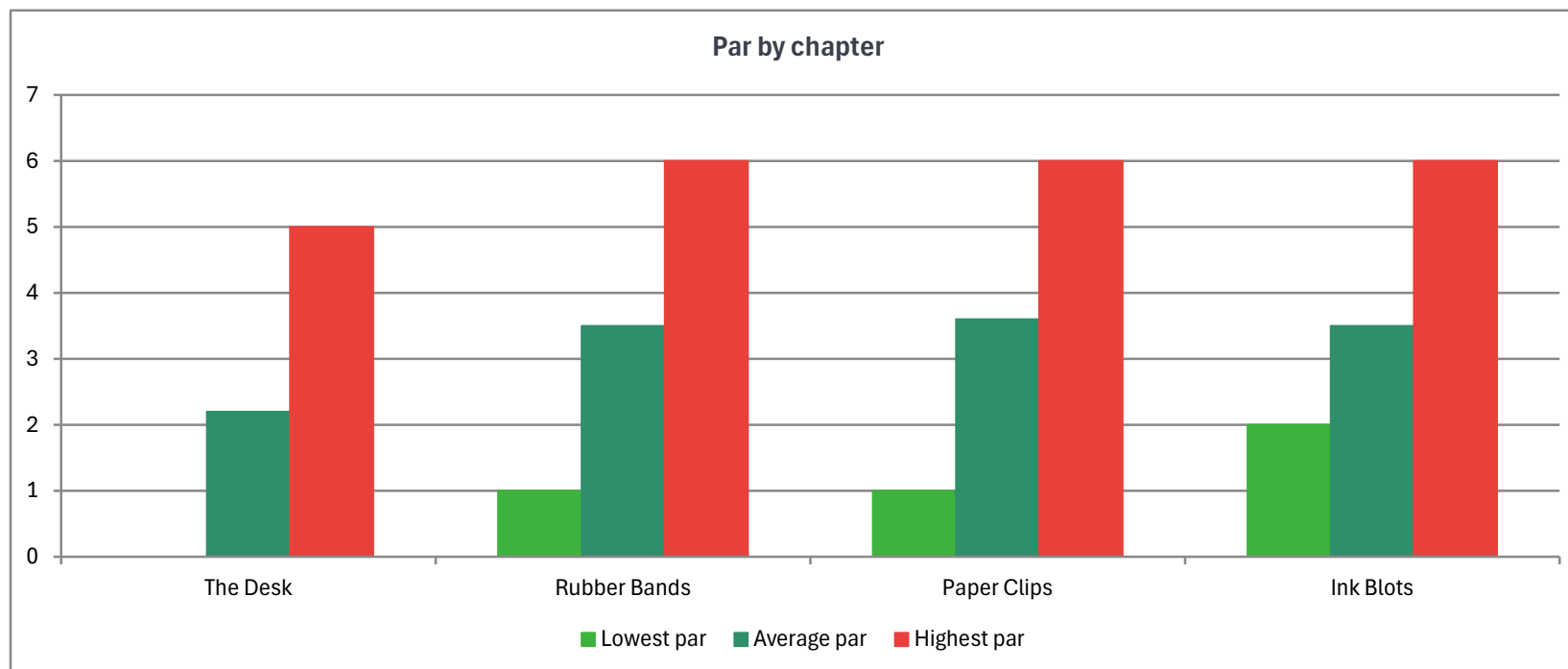
Formulas over the Levels sheet.

Chapter	Name	Levels	Average par	Lowest par	Highest par	Average pencils	Timed levels	Levels that need the full tray	Average opening odds	Greedy bot: average dead ends	Greedy bot: average 3-star share
1	The Desk	10	2.2	0	5	9.6	1	1	74%	15%	37%
2	Rubber Bands	10	3.5	1	6	11.7	2	1	71%	32%	16%
3	Paper Clips	10	3.6	1	6	12.3	2	3	100%	30%	49%
4	Ink Blots	10	3.5	2	6	12.0	2	3	83%	33%	29%

What moves together (Pearson correlation over the 40 levels)

Pair	r
Par and the greedy bot's dead ends	0.81
Tray needed and the greedy bot's dead ends	0.86
Pencils and the greedy bot's dead ends	0.75
Par and the opening odds	0.12
Pencils blocked at start and par	0.80

How to read it: par is the designed difficulty and the tray needed is how little room for error is left; the bots show where a careless player gets stuck. In Pull-out a careless player doesn't wander, they fill the tray: a level that needs all five slots punishes a single wasted park, whatever its par. Those are the levels to watch first in playtests.



What the measurements changed

Pull-out has not been playtested by people yet; every finding here comes from the solver and the bots.

Finding	What the measurement showed	What changed
Clips and blots tied to boxes did nothing	First version: a clip held a pencil until a number of boxes had rolled away, a blot until its colour's box had. Tools/probe.py: in 300 random deals per layout neither ever changed a par or the tray needed. The box order is fixed, so what they hold back is either needed before it is let go (the level can't be won) or it isn't (it costs nothing).	Released by the player's own pulls instead: a clip after that many pulls, a blot once a pencil of its colour is out. The par then changed in 9 of 85 and 7 of 48 winnable deals (clips) and 14 of 73 and 8 of 38 (blots), and the tray needed about as often.
Par alone was the wrong measure	A delay rarely makes a player park more; it makes more pencils wait at once. Rubber bands changed the tray needed in 30 of 48 winnable deals. In the shipped levels, 1 of 30 levels with a focus keep the same par without it and differ only in the tray needed (level 12).	The generator's focus check counts the tray a level needs (the smallest that still wins) as well as its par; a test checks it on every level with a focus.
A banded pull can park two	The C# solver priced a banded pull by its first pencil only, and both solvers used a search that assumes every step costs nothing or one. The EditMode tests caught the first: level 11's par was 1 in C#, 2 in Python.	Both solvers now search by cost with one queue per cost (Dial's algorithm). Level 14's par had been one too high and the level was dealt again.
The tray is the wall	The greedy bot dead-ends in at least 90% of its runs on levels 17 (98%), 20 (92%), 27 (94%), 34 (99%), 37 (96%); 4 of these 5 need the full tray. All the levels that need it: 10, 17, 26, 27, 30, 34, 37, 39.	Kept, as chapter peaks that reward reading the queue; the game names a dead end and pulses restart, and the sixth-slot booster rescues one. First on the watch list for playtests.
A new rule arrives on an easy level	Clockwork's playtest: a rule introduced on a hard level read as a difficulty spike. Here the first band is on level 11 (par 2), the first clip on 21 (par 2), the first blot on 31 (par 2); the chapters peak at par 5, 6, 6 and 6.	Each new mechanic gets a one-line tip on the level where it first appears; level 1 shows a hand on the first pencil to tap.
Still open: the clock and the boosters	Neither bot plays against the clock, so 3 s a pencil is not measured here; nor is whether three of each booster, plus one a chapter, is too generous.	Decide after the first playtest, from the time_up and booster_used events (Ad Test and KPI Plan).