Tuning minimal generalisations on a morphological learning task

Rácz, Péter 22 March, 2024

A cognitively plausible model of people's morphophonological intuitions in a Wug task is the Minimal Generalisation Learner (MGL). The MGL looks for generalisations of the type `CAD ~ CBD`, which can be expressed as a rule `A -> B` with the structural description of `C _ D`. We can train the MGL on existing words and it will predict people's responses to nonword stimuli. Here, we explore how the MGL can include short-term learning based on nonword stimuli. We use the data collected by Rácz, Beckner, Hay & Pierrehumbert (2020), who ran a baseline Wug task and a morphological learning task (which they call the ESP task), using an artificial coplayer. Both tasks focussed on the regular / irregular variation of the English past tense.

The baseline experiment

Nonwords

Rácz, Beckner, Hay & Pierrehumbert (2020) generated nonword verbs across four regular/irregular categories:

- drove ([aI]/[i] → [oU])
- sang ([I] → [ae])
- kept ([i] → [E]Ct)
- burnt ([3]/[E]/[I] → [3]/[E]/[I]Ct)

Nonwords were transcribed into the DISC phonetic alphabet. Examples are in Table 1.

burnt	drove	kept	sang
prill, [prIl]	fide, [f2d]	dreep, [drip]	chim, [JIm]
skrurn, [skr3n]	thride, [Tr2d]	streel, [stril]	thrim, [TrIm]
vrill, [vrIl]	squine, [skw2n]	squeep, [skwip]	smink, [smINk]
drurn, [dr3n]	brive, [br2v]	schmeem, [Smim]	frim, [frIm]
trurn, [tr3n]	sline, [sl2n]	shreep, [Srip]	quink, [kwINk]

1. Nonword examples.

Test data

202 participants, recruited on AMT, responded to the orthographic present tense form of each nonword in a simple carrier sentence in a forced-choice task. They could pick the regular or the irregular past tense form for each nonword, displayed on buttons. The regular past tense form was the -ed form. The irregular form depended on the verb class, as seen in Table 2.

category	word	regular_form	irregular_form
burnt	drell	drelled	drelt
burnt	sprell	sprelled	sprelt
drove	shride	shrided	shrode
drove	dwide	dwided	dwode
kept	sneep	sneeped	snept
kept	theep	theeped	thept
sang	schmim	schmimmed	schmam
sang	pring	pringed	prang

2. Regular and irregular choices in the Wug task.

Minimal Generalisation Learner (MGL)

Rácz, Beckner, Hay and Pierrehumbert (2020) trained the Minimal Generalisation Learner (MGL) on English verbs in CELEX and used it to make predictions for the nonwords. They trained the MGL on regular and irregular English verbs with a minimum frequency cutoff of 10: 4160 past/present verb transcriptions. They used the best parameters identified by Albright & Hayes (2003) for a similar task: lower and upper confidence limits of 55% and 95%.

The MGL generates 61 rules for the 156 target forms from the training data. Such rules have a structural description that matches a target nonword in the task and generates an output which is available to participants to pick. A rule that generates the `sing -> sang` pattern matches target forms for nonwords that look like `sing`. It generates one of the past tense forms available in the forced-choice task. A rule that generates the `sing -> sung` pattern does not generate an available past tense form.

rule	type	scope	hits	reliability	confidence
[] -> d [3, @, a]n __	regular	135	133	0.99	0.98
[] -> d [3, D, S, T, Z, l, r, s, z] __	regular	1443	1414	0.98	0.98
[] -> d [D, S, T, Z, n, s, z] __	regular	902	883	0.98	0.98
[] -> d [2, 4, 6, e, i, o, u]m __	regular	63	62	0.98	0.97
[] -> d [D, S, T, Z, f, s, v, z] __	regular	712	698	0.98	0.97

rule	type	scope	hits	reliability	confidence
[] -> d [b, m] __	regular	169	164	0.97	0.97
[] -> d [b, p] __	regular	214	207	0.97	0.96
[] -> d [J, S, T, s, t] __	regular	1364	1314	0.96	0.96
[] -> @d [D, J, S, T, Z, __, b, d, f, g, k, p, s, t, v, z, ~]»2d __	regular	13	13	1.00	0.96
[] -> @d [D, T, d, n, s, t, z]»2t __	regular	11	11	1.00	0.95
[] -> d [3, D, J, S, T, Z, __, d, l, n, r, s, t, z] __	regular	3183	3046	0.96	0.94
[] -> @d t __	regular	959	913	0.95	0.94
[] -> t [J, S, T, f, k, p, s, t, ~] __	regular	1779	1695	0.95	0.94
[] -> t [D, J, S, T, Z, __, d, s, t, z] __	regular	2020	1919	0.95	0.94
[] -> d [D, J, S, T, Z, __, b, d, f, g, k, p, s, t, v, z, ~] __	regular	2602	2458	0.94	0.94
[] -> d [D, J, N, S, T, Z, __, b, d, f, g, h, k, l, m, n, p, s, t, v, z, ~] __	regular	3510	3317	0.95	0.93
[] -> @d [2, 3, 4, 6, @, A, E, Q, V, a, e, o, {, »]d __	regular	99	89	0.90	0.89
2 -> o [3, D, J, S, T, Z, __, d, l, n, r, s, t, z]r» __ [d, t]	irregular	4	4	1.00	0.88
ip -> Ept [N, g, j, k, w]» __	irregular	3	3	1.00	0.85
ip -> Ept [j, l, r, w]» __	irregular	7	6	0.86	0.79
2 -> o r» __ [d, t]	irregular	9	7	0.78	0.73
2 -> o [S, Z, r]» __ [d, n, t]	irregular	13	9	0.69	0.66
2 -> o [2, 3, 4, 6, @, A, D, E, I, N, Q, U, V, Z, __, a, b, d, e, g, i, j, l, m, n, o, r, u, v, w, z, {, »]r» __ [D, Z, v, z]	irregular	3	2	0.67	0.59
I -> { [D, S, T, Z, l, r, s, z]» __ N	irregular	3	2	0.67	0.59
I -> { [D, J, S, T, Z, __, d, s, t, z]r» __ Nk	irregular	3	2	0.67	0.59
ip -> Ept [D, J, N, S, T, Z, __, b, d, f, g, h, j, k, l, m, n, p, r, s, t, v, w, z, ~]» __	irregular	12	7	0.58	0.56
[] -> t [D, N, Z, __, b, d, g, l, m, n, v, z]»3n __	irregular	4	2	0.50	0.47
il -> Elt [D, N, S, T, Z, f, m, n, s, v, z]» __	irregular	4	2	0.50	0.47
il -> Elt [D, J, N, S, T, Z, __, b, d, f, g, k, m, n, p, s, t, v, z, ~]» __	irregular	7	3	0.43	0.41
I -> { [N, g, j, w]» __ [N, m, n]	irregular	5	2	0.40	0.39
2 -> o [D, Z, __, d, l, n, r, z]» __ v	irregular	8	3	0.38	0.37
I -> { r» __ N	irregular	6	2	0.33	0.33
I -> { [D, J, S, T, Z, __, b, d, f, g, k, p, s, t, v, z, ~]r» __ N	irregular	9	3	0.33	0.33
2 -> o [D, Z, __, d, l, n, r, z]» __ [D, J, S, T, Z, __, b, d, f, g, k, p, s, t, v, z, ~]	irregular	12	4	0.33	0.33
I -> { [D, J, S, T, Z, __, d, l, n, r, s, t, z]» __ Nk	irregular	12	4	0.33	0.33
I -> { [D, S, T, Z, l, r, s, z]» __ N	irregular	10	3	0.30	0.30
[] -> t [m, w]»El __	irregular	7	2	0.29	0.29
2 -> o [N, j, l, m, n, r, w]» __ t	irregular	14	4	0.29	0.28

rule	type	scope	hits	reliability	confidence
2 -> o [D, Z, __, d, l, n, r, z]» __ [D, Z, __, b, d, g, v, z]	irregular	34	9	0.26	0.26
2 -> o [N, j, l, m, n, r, w]» __ [d, t]	irregular	22	8	0.36	0.26
I -> { [b, m, v, w]» __ [N, __, b, d, g, m, n]	irregular	8	2	0.25	0.26
2 -> o [D, J, S, T, Z, __, d, s, t, z]» __ [D, N, Z, m, n, v, z]	irregular	9	2	0.22	0.23
2 -> o [D, J, S, T, Z, __, d, l, n, r, s, t, z]» __ [D, N, Z, __, b, d, g, m, n, v, z]	irregular	18	4	0.22	0.22
I -> { [D, S, T, Z, l, r, s, z]» __ [J, N, __, b, d, g, k, m, n, p, t, ~]	irregular	15	3	0.20	0.21
I -> { [j, r, w]» __ [N, m, n]	irregular	15	3	0.20	0.21
I -> { [D, J, S, T, Z, __, d, l, n, r, s, t, z]» __ N	irregular	35	7	0.20	0.20
I -> { [D, S, T, Z, f, h, j, l, r, s, v, w, z]» __ [N, m, n]	irregular	22	4	0.18	0.18
I -> { [D, J, S, T, Z, __, b, d, f, g, k, p, s, t, v, z, ~]» __ [N, m, n]	irregular	13	2	0.15	0.16
2 -> o [D, J, S, T, Z, __, d, l, n, r, s, t, z]» __ [D, J, N, S, T, Z, __, b, d, f, g, k, m, n, p, s, t, v, z, ~]	irregular	29	5	0.17	0.16
2 -> o [D, J, S, T, Z, __, d, l, n, r, s, t, z]» __ [D, N, Z, m, n, v, z]	irregular	44	7	0.16	0.14
i -> o [b, f, m, p, v, w]» __ [D, J, S, T, Z, __, b, d, f, g, k, p, s, t, v, z, ~]	irregular	30	4	0.13	0.14
I -> { [j, r, w]» __ [N, m, n]	irregular	45	5	0.11	0.10
[] -> t [E, I]l __	irregular	46	4	0.09	0.09
2 -> o [D, N, S, T, Z, f, m, n, s, v, z]» __ [d, n, t]	irregular	36	3	0.08	0.09
i -> o [D, J, N, S, T, Z, __, b, d, f, g, h, j, k, l, m, n, p, r, s, t, v, w, z, ~]» __ [D, Z, l, v, z]	irregular	50	4	0.08	0.08
I -> { [D, J, S, T, Z, __, b, d, f, g, k, p, s, t, v, z, ~]» __ [N, m, n]	irregular	58	4	0.07	0.07
I -> { [D, N, Z, __, b, d, g, j, l, m, n, r, v, w, z]» __ [N, m, n]	irregular	83	6	0.07	0.07
2 -> o [D, N, S, T, Z, f, h, j, l, m, n, r, s, v, w, z]» __ [d, n, t]	irregular	66	10	0.15	0.07
2 -> o [D, J, N, S, T, Z, __, b, d, f, g, k, m, n, p, s, t, v, z, ~]» __ [D, J, N, S, T, Z, __, b, d, f, g, k, m, n, p, s, t, v, z, ~]	irregular	75	4	0.05	0.06
[] -> t [d, n, t] __	irregular	1545	1144	0.74	0.02

3. Rules from Celex.

Rules take the structural description of input -> output / context. Multiple rules can apply to the same input form. For the majority of forms, there is one regular and one irregular rule available. For some, there is no irregular rule. For some, there are more regular rules. Examples can be seen in Table 4.

category	word	regular	irregular
burnt	skell	1	1
drove	squine	1	1
drove	chite	2	1
drove	yide	3	1
kept	skeep	1	1
sang	thring	1	1
sang	grink	2	1

Table 4. Possible number of regular / irregular rules for some forms.

Following both Rácz, Beckner, Hay & Pierrehumbert (2020) and Albright & Hayes (2003) we can pick the best regular and the best irregular rule and calculate a weight, which is the confidence of the best regular rule / (the confidence of the best regular rule + the confidence of the best irregular rule). If there is no irregular rule, this will default to 1. We will work with the rules that are best rules for any form and call these the relevant rules.

Results

Figure 1. Baseline task: Word ratings and MGL predictions

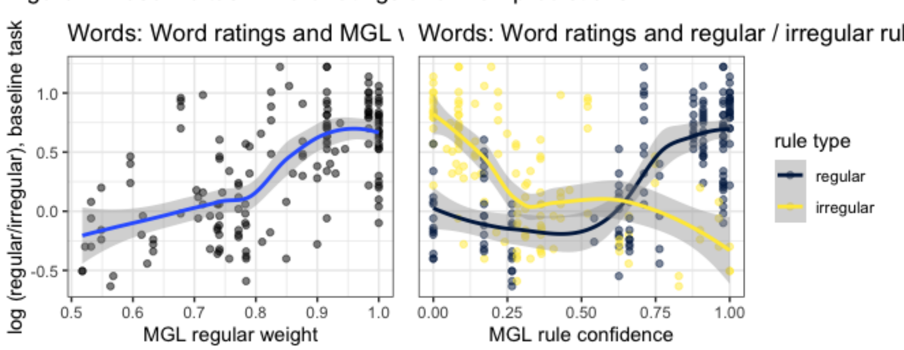


Figure 1 shows how MGL predictions correlate with participant responses in the baseline etask. The left panel shows the relationship between word weights (x axis) and the log odds of regular and irregular choices made by participants in the baseline task (y axis). The right panel breaks down the weight into its two components: the best regular and irregular rule for each word. Since, on the whole, regular rules have higher confidence than irregular rules, we rescaled rule confidence across these groups so they are more directly comparable.

The MGL weights correlate with the response odds. We see that the trajectory of this relationship is built up from two opposite trajectories for best rules. For words with low regular weight, this weight comes from irregular rules that have low confidence themselves but, relatively speaking, outweigh the relevant regular rules. For words with higher regular weight, this comes from two things. First, high-confidence, large-scale regular rules apply to these, and these rules bring up the regular weight. Second, the relevant irregular rules have very low confidence. This breakdown of the MGL's regular weight will become relevant later.

Note that, from the MGL's perspective, the main technical difference between regular and irregular rules is that regular rules tend to have broader structural descriptions and will fit more existing words. Whether a rule generates an `-ed` ending or e.g. changes a vowel is not structurally different from the model's point of view.

Rácz, Beckner, Hay & Pierrehumbert (2020) and Albright & Hayes (2003) likewise find that the minimal generalisations (rules) of the MGL are more accurate in predicting participant responses than an instance-based learner, despite the higher level of abstraction.

The ESP experiment

Rácz, Beckner, Hay & Pierrehumbert (2020) ran a second online experiment using new participants and the nonwords from the baseline experiment. Each participant went through three blocks. First, in the pretest phase, they responded to 52 standalone nonwords in a forced-choice task, identical to the baseline experiment. Second, in the ESP test phase, they responded to a new set of 52 nonwords. This time they were playing against a co-player and had to guess the coplayer's pick in each trial. Correct guesses were rewarded with a point. Coplayer behaviour was based on the participant's specific pretest behaviour and the baseline data. Third, in the posttest phase, they responded to a new set of 52 nonwords, playing alone again. The ESP design is used widely in tasks where it is important for descriptions to match, like image tagging. ESP refers to the fact that participans have to "read" each other's minds to converge on a description. In this particular case, the participant had to do all the mindreading, since the coplayer's choices were set in advance.

Coplayers varied across two conditions. In terms of (A) rate of regularisation, the coplayer had (i) the same regularisation rate as the participant in the pretest, (ii) regularised 40% more verbs, (iii) regularised 40% fewer verbs. Participants who regularised too much or too little (so that the entire effect of this shift would have been capped by the floor or the ceiling of the 52 verbs in the ESP test) were excluded. In terms of (B) lexical distribution, the coplayer regularised the first n% verbs (n depending on A) that were rated most regular in the baseline task (the typical coplayer), the first n% verbs that were rated most irregular in the baseline task (the reversed coplayer), or n% verbs at random (the random coplayer). This means that a typical coplayer makes choices that are characteristic of an average participant. A reversed coplayer turns these choices upside down.

Results

Rácz, Beckner, Hay & Pierrehumbert (2020) found that the coplayer changed participant behaviour. Nonword ratings shifted from the pretest to the posttest. If participants rated words as highly regular in the pretest, they rated these more irregular in the posttest, after interacting with the reversed versus the typical coplayer. Since no participant saw the same verb twice, this effect was due to lexical, rather than word priming. The reversed coplayer used verbs in a certain way, and the participant rated similar verbs in the posttest in a certain way. The difference with the random coplayer was less clearcut. This can be seen in Figure 2.

Figure 2. ESP task: Word ratings in pretest and posttest

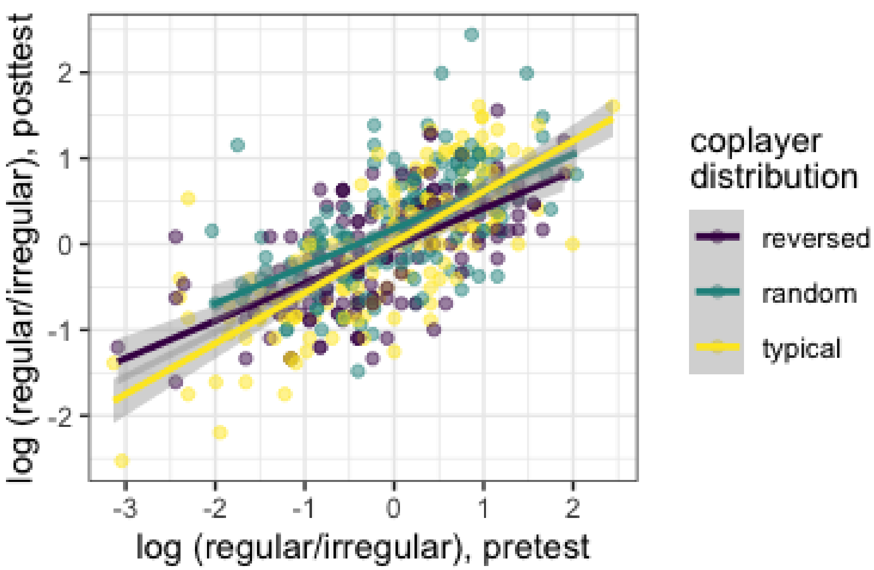


Figure 2 compares response log odds in the pretest and the posttest. These are correlated: participants make similar choices at the beginning of the experiment and at the end, after encountering the coplayer. However, this correlation is weaker if participants played a reversed coplayer, demonstrating the coplayer's influence.

What changes in the posttest

The MGL predicts participant responses in general. So, if we look at how MGL prediction accuracy varies across coplayers in the posttest, we find a pattern similar to how these responses shift in the posttest.

The main effect of interest is coplayer lexical distribution. The MGL weight has a stronger effect on posttest responses if the participant played a typical coplayer. This makes sense: a typical coplayer reinforces existing lexical distributions. The MGL's predictive power diminishes when it is set against participants who met a reversed coplayer. What are the mechanics of this shift?

Figure 3. ESP posttest: Word ratings in posttest
and MGL weights.

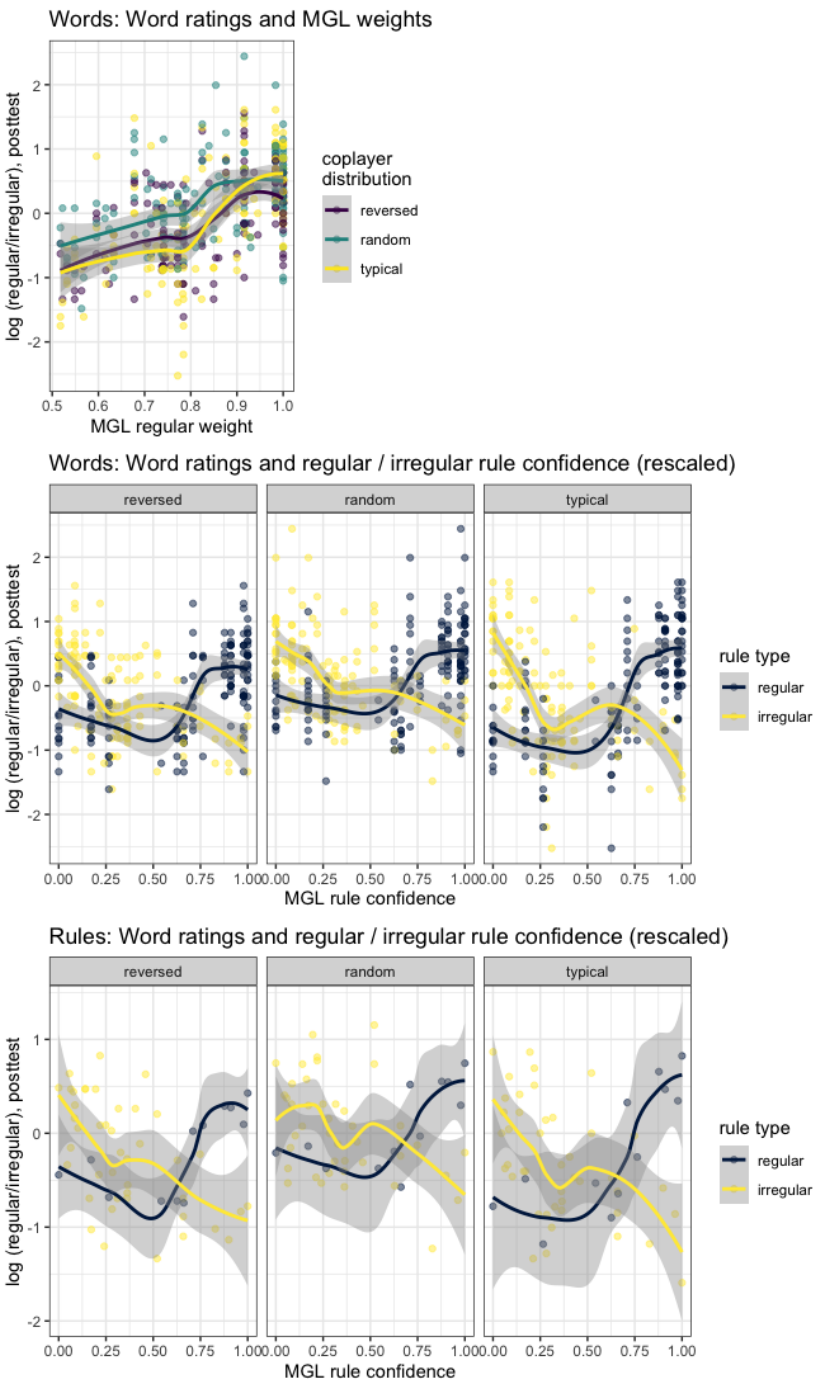
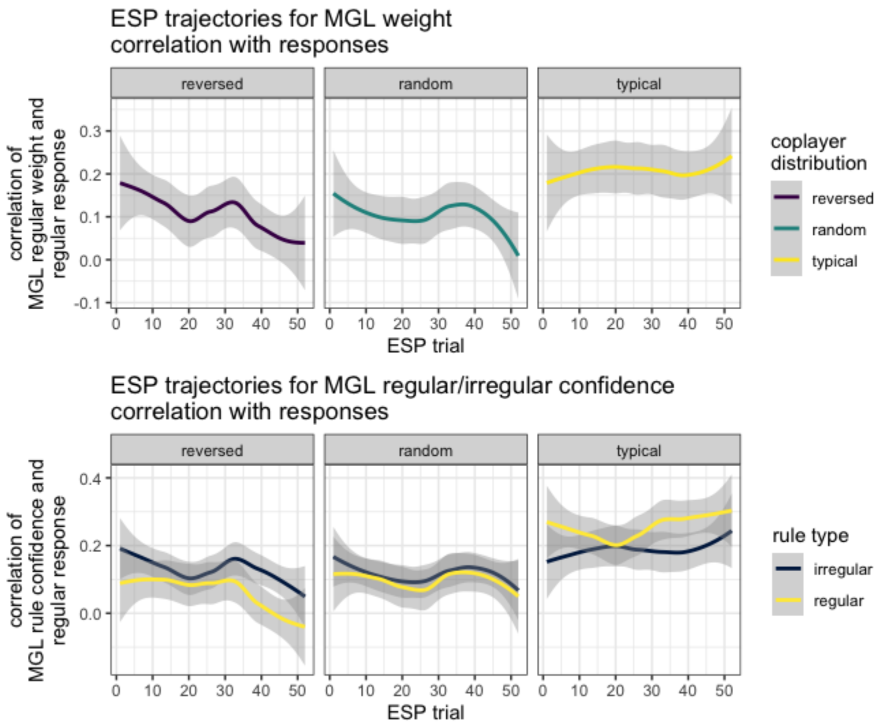


Figure 3 shows the correlation of MGL predictions with participant responses in the posttest, split across coplayer lexical distribution.

The top panel shows individual word weights. It is similar to Figure 2. The MGL is the best at predicting the posttest condition that is closest to the baseline, which is the typical condition. The correlation is weaker for responses by participants who encountered a reversed or a random coplayer. The middle panel shows this relationship broken down to the two contributing factors to an MGL weight: the best regular and the best irregular rule for each word. Looking at Figure 2, we said that low weights follow from relevant irregular rules outweighing regular rules in confidence, while, for high weights, this relationship is reversed. Here we see that participants follow this pattern in the typical condition but diverge from it in the reversed and random conditions: their choices reflect a smaller difference between regular and irregular rules than what is predicted by the MGL, and this means that the MGL undershoots irregular verbs and overshoots regular verbs. We see the same relationship in the bottom panel, where we look at confidences and ratings aggregated over individual rules rather than individual words. For each rule, we count the regular and irregular posttest responses for all the nonword verbs in its scope, given that, for each verb, no rule of higher confidence was available. These are the verbs for which this is the best regular / irregular rule. We then plot this against the rule's confidence. Participants in the random and reversed conditions act as if the regular rules and the irregular rules were closer to each other, which is why the MGL has lower accuracy than in the typical condition (or the baseline task).

Posttest patterns come to existence during the interactive session with the coplayer, the ESP task. Participants gradually diverge from the MGL predictions during the ESP task. We visualise this in Figure 4. The top panel shows the relationship between MGL weights and participant responses in each of the 52 trials of the ESP task. For each trial, we calculated a Pearson correlation between word weights and participant responses across the three coplayer lexical distributions. Trials are shown on the x axis. The correlations are shown on the y axis. The correlations vary a lot, so we only plot a loess smooth for each condition. We see that the correlations hold steady for the typical condition, where the coplayer makes lexically typical choices. They gradually deteriorate across the other conditions, in which the coplayer's choices go against the participants' (and the MGL's) expected lexical distributions. The bottom panel breaks this down into regular and irregular rules. The two rule types move in tandem: it is not the case that the overall trajectory shifts because of the behaviour of the regular rules, for instance.

Figure 4. Rule weight and confidence across ESP.



Modelling

The main result of Rácz, Beckner, Hay & Pierrehumbert (2020) is that, when you expose participants to a lexical distribution in the ESP task, they will extend this distribution to previously unseen forms in the posttest, to some extent.

The MGL can capture this shift through rules or minimal generalisations. Participants see different verbs in the ESP task and the posttest. The rules that apply to these verbs will overlap. This will be especially true for rules that have broad structural descriptions and thus apply to many forms. These tend to be regular rules.

rule	type	scope	phase
[] -> d [D, J, N, S, T, Z, __, b, d, f, g, h, k, l, m, n, p, s, t, v, z, ~] __	regular	3510	esp, posttest
[] -> d [3, D, J, S, T, Z, __, d, l, n, r, s, t, z] __	regular	3183	esp, posttest

rule	type	scope	phase
[] -> t [J, S, T, f, k, p, s, t, ~] __	regular	1779	esp, posttest
[] -> t [d, n, t] __	irregular	1545	esp, posttest
[] -> d [3, D, S, T, Z, l, r, s, z] __	regular	1443	esp, posttest
[] -> d [J, S, T, s, t] __	regular	1364	esp, posttest
[] -> d [D, S, T, Z, n, s, z] __	regular	902	esp, posttest
[] -> d [D, S, T, Z, f, s, v, z] __	regular	712	esp, posttest
[] -> d [b, p] __	regular	214	esp, posttest
[] -> d [b, m] __	regular	169	esp, posttest
[] -> d [3, @, a]n __	regular	135	esp, posttest
[] -> d [2, 4, 6, e, i, o, u]m __	regular	63	esp, posttest
[] -> t [E, I]l __	irregular	46	esp, posttest
I -> { [j, r, w]» __ [N, m, n]	irregular	45	esp, posttest
2 -> o [D, J, S, T, Z, __, d, l, n, r, s, t, z]» __ [D, N, Z, m, n, v, z]	irregular	44	esp, posttest
2 -> o [D, Z, __, d, l, n, r, z]» __ [D, Z, __, b, d, g, v, z]	irregular	34	esp, posttest
2 -> o [N, j, l, m, n, r, w]» __ [d, t]	irregular	22	esp, posttest
2 -> o [N, j, l, m, n, r, w]» __ t	irregular	14	esp, posttest
2 -> o [S, Z, r]» __ [d, n, t]	irregular	13	esp, posttest
I -> { [D, S, T, Z, l, r, s, z]» __ N	irregular	10	esp, posttest
2 -> o r» __ [d, t]	irregular	9	esp, posttest
I -> { [b, m, v, w]» __ [N, __, b, d, g, m, n]	irregular	8	esp, posttest
I -> { [N, g, j, w]» __ [N, m, n]	irregular	5	esp, posttest
2 -> o [3, D, J, S, T, Z, __, d, l, n, r, s, t, z]r» __ [d, t]	irregular	4	esp, posttest

rule	type	scope	phase
i -> o [D, J, N, S, T, Z, __, b, d, f, g, h, j, k, l, m, n, p, r, s, t, v, w, z, ~]» __ [D, Z, l, v, z]	irregular	50	esp
I -> { [D, J, S, T, Z, __, d, l, n, r, s, t, z]» __ N	irregular	35	esp
I -> { [D, S, T, Z, l, r, s, z]» __ [J, N, __, b, d, g, k, m, n, p, t, ~]	irregular	15	esp
I -> { [j, r, w]» __ [N, m, n]	irregular	15	esp
I -> { [D, J, S, T, Z, __, d, l, n, r, s, t, z]» __ Nk	irregular	12	esp
ip -> Ept [D, J, N, S, T, Z, __, b, d, f, g, h, j, k, l, m, n, p, r, s, t, v, w, z, ~]» __	irregular	12	esp
2 -> o [D, J, S, T, Z, __, d, s, t, z]» __ [D, N, Z, m, n, v, z]	irregular	9	esp
2 -> o [D, Z, __, d, l, n, r, z]» __ v	irregular	8	esp
[] -> t [m, w]»El __	irregular	7	esp
il -> Elt [D, J, N, S, T, Z, __, b, d, f, g, k, m, n, p, s, t, v, z, ~]» __	irregular	7	esp
I -> { r» __ N	irregular	6	esp
I -> { [D, N, Z, __, b, d, g, j, l, m, n, r, v, w, z]» __ [N, m, n]	irregular	83	posttest
2 -> o [D, N, S, T, Z, f, h, j, l, m, n, r, s, v, w, z]» __ [d, n, t]	irregular	66	posttest
I -> { [D, J, S, T, Z, __, b, d, f, g, k, p, s, t, v, z, ~]» __ [N, m, n]	irregular	58	posttest
2 -> o [D, N, S, T, Z, f, m, n, s, v, z]» __ [d, n, t]	irregular	36	posttest
i -> o [b, f, m, p, v, w]» __ [D, J, S, T, Z, __, b, d, f, g, k, p, s, t, v, z, ~]	irregular	30	posttest
I -> { [D, S, T, Z, f, h, j, l, r, s, v, w, z]» __ [N, m, n]	irregular	22	posttest
2 -> o [D, J, S, T, Z, __, d, l, n, r, s, t, z]» __ [D, N, Z, __, b, d, g, m, n, v, z]	irregular	18	posttest
[] -> @d [D, J, S, T, Z, __, b, d, f, g, k, p, s, t, v, z, ~]»2d __	regular	13	posttest
I -> { [D, J, S, T, Z, __, b, d, f, g, k, p, s, t, v, z, ~]r» __ N	irregular	9	posttest
ip -> Ept [j, l, r, w]» __	irregular	7	posttest
[] -> t [D, N, Z, __, b, d, g, l, m, n, v, z]»3n __	irregular	4	posttest
ip -> Ept [N, g, j, k, w]» __	irregular	3	posttest

Table 5. Rule overlaps for one participant

To illustrate this point, we show one participant in Table 5. Taken together, 47 relevant rules apply to the 104 verbs that our participant sees in the ESP task and the posttest. 24 rules overlap: they apply to some verbs in both tasks. 11 rules only apply to verbs in the ESP task. Following the MGL logic, whatever the participant learned about these verbs won't carry over to the posttest. 12 rules only apply to verbs in the posttest. The participant didn't learn anything new about these verbs. Unsurprisingly, the rules that overlap have much larger scopes (a mean of 636) than those which do not (a mean of 23). 11 overlapping rules are regular, only 1 non-overlapping rule is regular.

This suggests that whatever the participant learns about the regular rules will be far more influential in the posttest than what they learn about irregular rules.

One way to express this learning is to adjust rule confidence for rules in the ESP task. We will do this in the following way: For each participant and rule in the ESP task, we tally the number of regular and irregular responses by the participant in the ESP task. We do this for the reversed condition only.

If the rule is regular, it works very well if every verb it applies gets a regular response. For every regular response, the rule is rewarded. For every irregular response, the rule incurs a penalty. If the rule is irregular, all the verbs in its scope should be irregular. For every irregular response, the rule is rewarded. For every regular response, the rule incurs a penalty.

These total rewards and penalties are used to update the rule's confidence. If the participant starts the ESP task with a very strong regular rule but keeps picking irregular forms for verbs in the rule's scope, the rule's confidence is steadily demoted. If the rule works well across the task, its confidence will increase. The rate of this increase / decrease is controlled by the parameter `learning rate`, which ranges between 0.05 and 1.5, with a .05 step, resulting in 90 fits of the learner. We fit the rule updater in three configurations: (i) updating both regular and irregular rules, (ii) updating regular rules only, (iii) updating irregular rules only.

Figure 5. Updated rule confidence and predictive accuracy in the posttest.

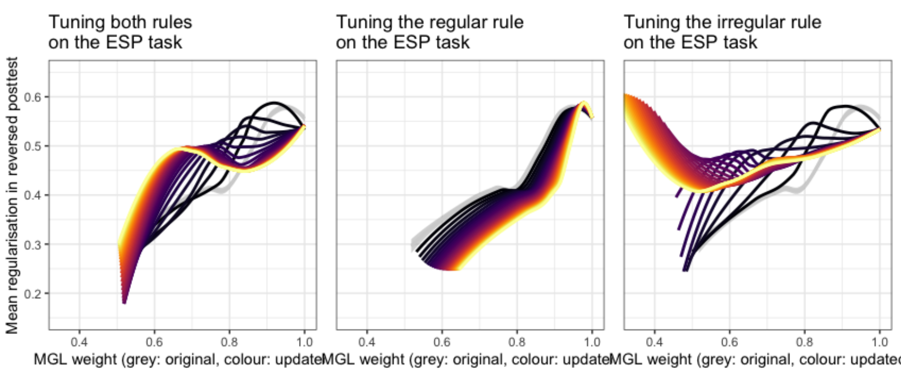


Figure 5 shows how the updated rules fare in the reversed condition posttest. The y axis shows mean participant regularisation, the x axis shows MGL regular weight for words in the posttest. We only show loess smooths of posttest regularisation and MGL weight. The grey trajectory is the accuracy of the original MGL's regular weights. Each coloured smooth shows weights from one updated

MGL. Colour shows the learning rate: darker colours mean that each ESP response updated the relevant rules to a small extent, lighter colours, to a large extent. The three panels show how the updated models change in accuracy if we update both rules (left), the regular rule only (middle), and the irregular rule only (right).

We used a simple Spearman correlation between mean word regular response and updated MGL weight to find the best model. The Spearman correlation between mean regular response and the original MGL model weights is 0.43. The best updated model only updates the regular rules and has a learning weight of 1.15. Its correlation coefficient is 0.59, which is a clear improvement on the original MGL.

Discussion

We trained the Minimal Generalisation Learner on real English verbs and tested it on results of a forced-choice Wug task in which participants had to pick the regular or irregular past tense form for nonwords. Like in earlier work (Albright & Hayes 2003), the MGL can predict participant behaviour in a Wug task.

We also tested the MGL on results from a morphological convergence experiment in which participants are exposed to coplayers and have to agree with them on choices in a forced-choice Wug task. Participants will converge to the lexical distributions of the coplayer and subsist with this distribution in a subsequent posttest, extending it to novel noword forms.

Since the MGL was trained on real verbs which comprise a typical lexical distribution, its predictions will be less accurate when measured against participant choices after participants have been exposed to a coplayer who reverses the typical distribution. We can tune the MGL to better fit participant responses by using data from the participant's interaction with the coplayer and updating each MGL rule based on participant choices during this interaction.

The morphological convergence experiment is unusual in how much it showcases English irregular inflectional morphology. In reality, the vast majority of existing English verb types is regular and irregular lexical gangs take on new members very infrequently. In the experiment, the use of irregular forms is rampant, both by the coplayers and the participants (the mean rate of use for regular forms was 41% in the baseline task).

One would then assume that if participants operate by establishing and updating minimal morphophonological generalisations, they would tackle the reversed coplayer by updating the irregular generalisations, establishing heuristics such as "if words look like <> they are much more likely to have a past tense like <> today". However, our simulations show that the best way to capture the change in participant behaviour is to update the regular rules only. This is likely because these rules apply to many forms: A regular rule adjusted by many verbs in the ESP task will apply to many verbs in the posttest. In contrast, while updating irregular rules may well better capture the shifts we see in participant behaviour, they are too fragmented for this to carry over from the learning phase (the ESP task) to retesting (the posttest). This would mean that the heuristic above should be reformulated as: “all words that look really regular should actually be less regular today”.