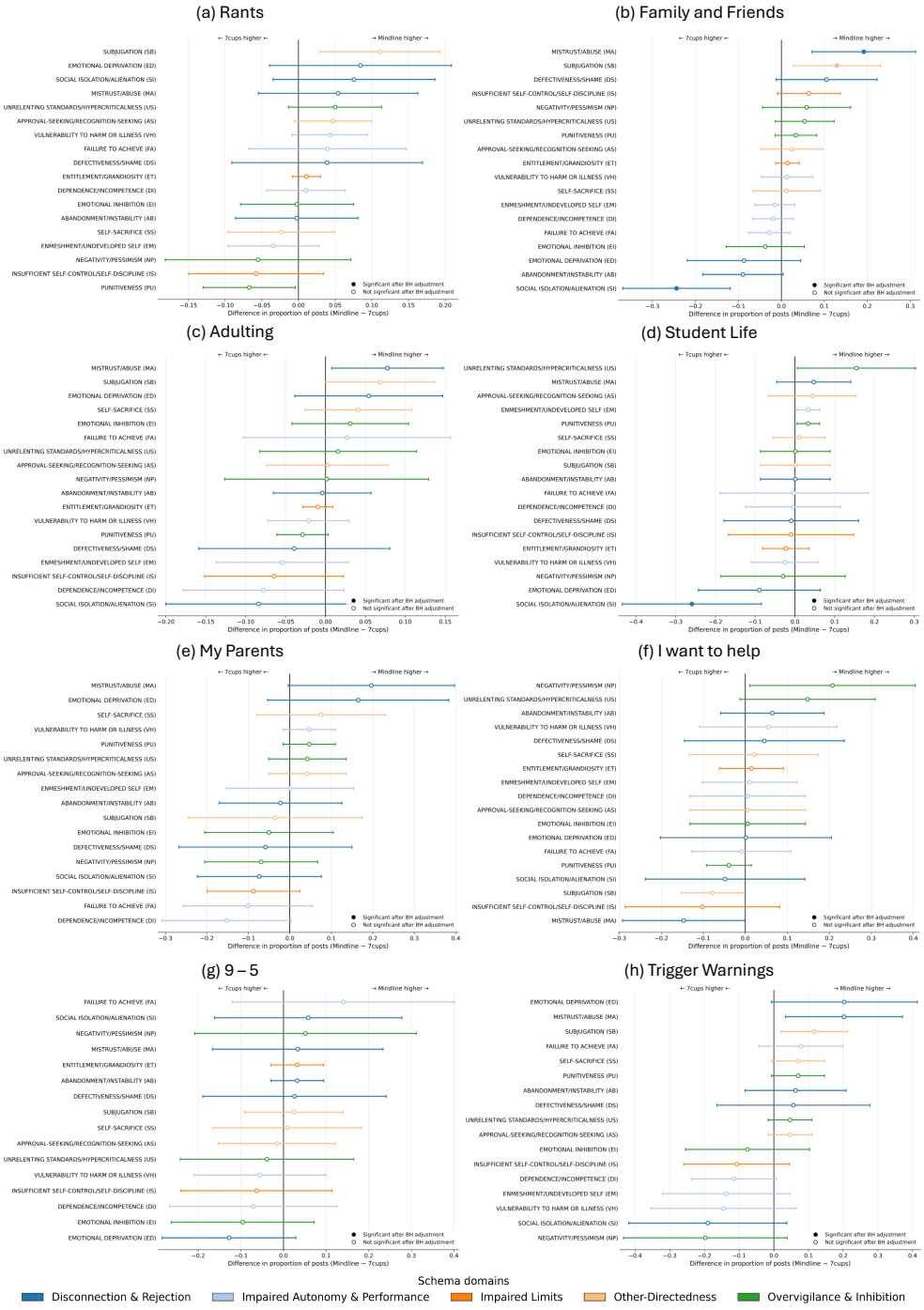
Supplementary materials for: "AI-aided Population-Level Comparisons of Early Maladaptive Schemas Between Singaporean and Global Online Communities"

Supplementary Section 1 Cross-platform EMS comparisons across topics

Note that here we present the differences in EMS prevalences for all topics, whereas in the manuscript we only presented the results for the 4 topics of (a) Rants, (b) Family and Friends, (c) Adulting and (d) Student Life. This is because the topics of (e) My parents, (f) I want to help, (g) 9–5 and (h) Trigger Warnings has (1) $n < 50$ samples for either mindline or 7cups, (2) majority of the EMS's within these topics do not meet the assumption of large-sample conditions for binomial approximations (i.e., $np \geq 5$ and $n(1-p) \geq 5$)[1] [1] and (3) reflected comparatively broad or generic discussion contexts (e.g., "9–5"), rather than more well-defined interpersonal or developmental domains such as Family and Friends or Student Life.

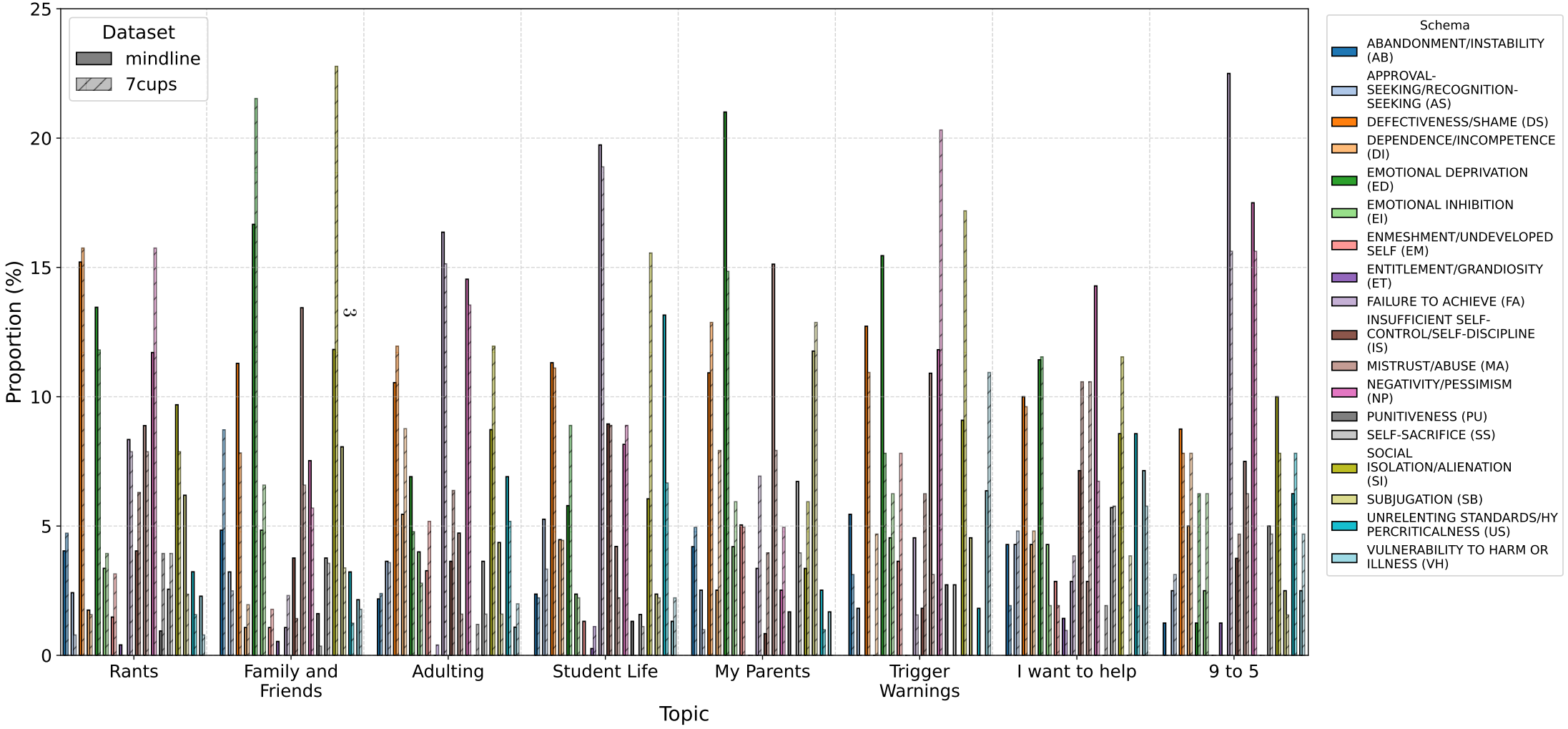
[1] These conditions are arbitrarily defined heuristic guidelines rather than strict thresholds.



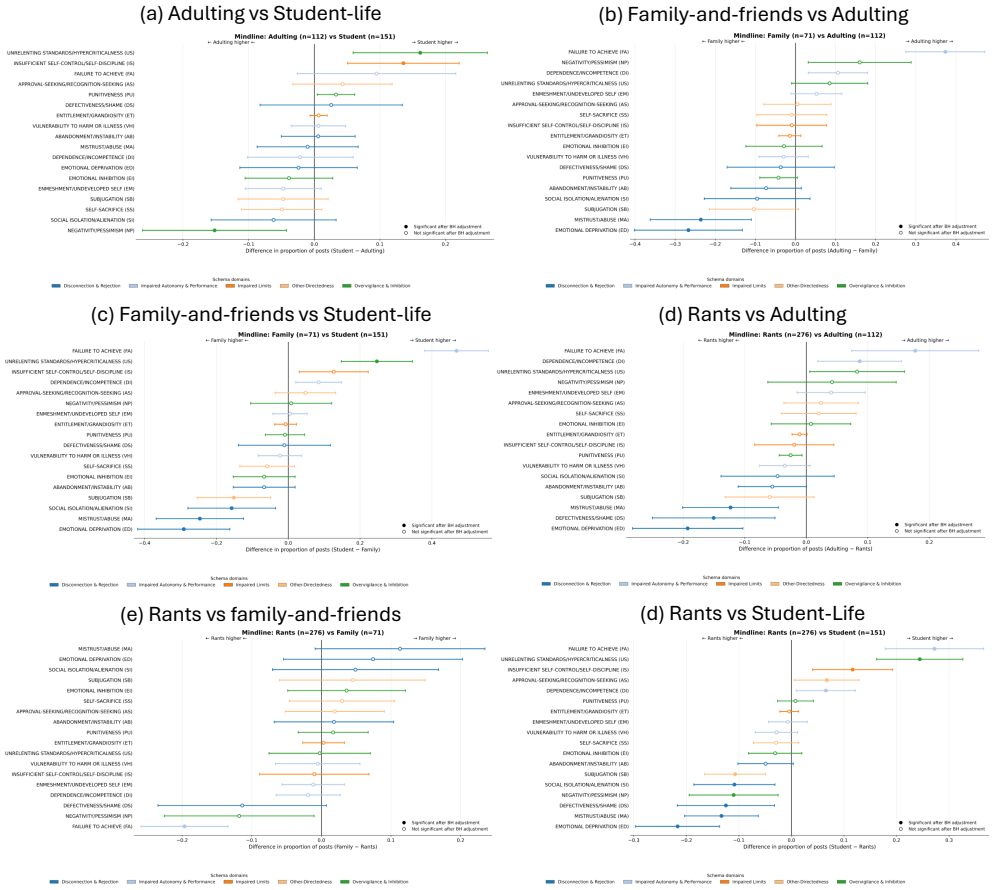
Supplementary Figure 1 Differences in schema prevalence between Mindline and 7cups across each topic

Forest plots illustrating the differences in schema prevalence between `Mindline` and `7cups` across topics spanning (a) Rants, (b) Family and Friends, (c) Adulting, (d) Student Life, (e) My parents, (f) I want to help, (g) 9–5 and (h) Trigger Warnings. Forest plots are represented in the form of 95% confidence intervals, with points representing effect estimates, filled points indicate statistical significance with Benjamini-Hochberg (BH) adjusted p-values at $\alpha = 0.05$, and color coded lines representing the distinct schema domains. Note that here we present the results of eight topics (as opposed to the main manuscripts four) as we included topics that did not necessarily fit the above-mentioned criteria, but contained sufficient samples to perform a cross-platform comparison. In addition, the topics of "Money Talk" and "Hobbies" were excluded as some EMS's were not even present (or had a sample size of 1) within these topics, rendering it impossible to perform a Z-test score of comparisons.

Supplementary Figure 2 Distribution of EMS between the Mindline and 7cups across topics.
An illustrative comparison of the distribution of EMS between both Mindline and 7cups when stratified by topics. Topics are represented across the x-axis, with schemas coded with distinct colors and the dataset distinguished by contrasting shades. Note that here we present the results of eight topics (as opposed to the main manuscripts four) as we included topics that did not necessarily fit the above-mentioned criteria, but contained sufficient samples to perform a cross-platform comparison. In addition, the topics of "Money Talk" and "Hobbies" were excluded as some EMS's were not even present (or had a sample size of 1) within these topics, rendering it impossible to perform a Z-test score of comparisons.



Supplementary Figure 3 Differences in schema prevalence between topics in Mindline Forest plots illustrating the differences in schema prevalence between each pair of topics in Mindline, in the following order: (a) Adulting vs Student-life, (b) Family-and-friends vs Adulting, (c) Family-and-Friends vs Student-life, (d) Rants vs Adulting, (e) Rants vs Family-and-Friends, and (d) Rants vs student life. Forest plots are represented in the form of 95% confidence intervals, with points representing effect estimates, filled points indicate statistical significance with Benjamini-Hochberg (BH) adjusted p-values at $\alpha = 0.05$, and color coded lines representing the distinct schema domains.



References

[1] W. G. Cochran, "Some methods for strengthening the common χ^2 tests," *Biometrics*, vol. 10, no. 4, pp. 417–451, 1954.