Appendix A

Supplementary Text

Methods

In addition to the structural equation modeling of learned control in the main text, we here present a 2 x 2 x 2 repeated-measures ANOVAs for biased (PC-90/10 and PS-80/20) and unbiased (PC-50 and PS-50) items, with Proportion Congruent (mostly incongruent/mostly congruent) or Switch (mostly task-repeat/mostly task-switch) and current trial Congruency (congruent/incongruent) or Type (task-repeat/task-switch) as within-participants factors and Block Order (MC-MC-MI-MI/MI-MI-MC-MC or MR-MR-MR-MR-MS-MS-MS-MS/MS-MS-MS-MS-MR-MR-MR-MR) as a between-participant factor. Block order was added to the analysis, because previous studies (Bejjani et al., 2020; Bejjani & Egner, 2021) have found that shifting from blocks with more control-demand (e.g., more conflict) to blocks with less control-demand (e.g., less conflict) results in a smaller congruency effect (or accuracy switch cost) across blocks than shifting from less to more control-demand. All the same filters that were described in the main text were applied to the analyses here.

Additionally, because we counterbalanced the order of PC and PS contexts across participants (cf. Bejjani et al.,2021; Bejjani & Egner, 2021) and the first two or four blocks were of one PC/PS context, we assessed whether participants showed control-learning effects early on in the task. We thus dropped PC or PS context as a within-subjects factor and treated block order as a between-subjects factor that reflected the difference in the PC or PS contexts.

See Supplementary Tables 1-8 (1-4 for within-participants PC/PS, 5-8 for between-participants PC/PS) for the full ANOVA results.

Results

Biased PC-90/10 and PS-80/20 Items

Reaction Time (ms). Within the LWPC, we observed a significant within-participant interaction between proportion congruency and congruency ($F(1,614) = 414.38$, $p < 0.001$, $\eta_p^2 = 0.40$): the congruency effect was reduced for PC-10 (MI first: 663 – 631 = 32; MC first: 643 – 616 = 27) compared to PC-90 (MI first: 689 – 603 = 86; MC first: 723 – 619 = 104) items, replicating the biased list/item effect. We also observed an interaction between block order (between-participants PC) and congruency ($F(1,731) = 285.62$, $p < 0.001$, $\eta_p^2 = 0.28$), with a smaller congruency effect for PC-10 (664 – 634 = 30) than PC-90 (722 – 620 = 122) items.

Within the LWPS, we observed a significant within-participant interaction between proportion switch and trial type ($F(1,950) = 194.41$, $p < 0.001$, $\eta_p^2 = 0.17$): the switch cost was reduced for PS-80 (MS first: 750 – 739 = 11; MR first: 723 – 705 = 18) compared to PS-20 (MS first: 726 – 700 = 26; MR first: 754 – 717 = 37) items, replicating the biased list/item effect. We also observed an interaction between block order (between-participants PS) and trial type ($F(1,953) = 120.68$, $p < 0.001$, $\eta_p^2 = 0.11$), with a smaller switch cost for PS-80 (749 – 738 = 11) than PS-20 (754 – 717 = 37) items.

Accuracy (%). Within the LWPC, we observed a significant within-participant interaction between proportion congruency and congruency ($F(1,955) = 829.80$, $p < 0.001$, $\eta_p^2 = 0.46$): the congruency effect was reduced for PC-10 (MI first: 85.40 – 87.91 = 2.51; MC first: 90.50 – 89.68 = 0.82) compared to PC-90 (MI first: 77.80 – 94.16 = 16.36; MC first: 70.41 – 92.70 = 22.29) items, replicating the biased list/item effect. We also observed an interaction between block order (between-participants PC) and congruency ($F(1,955) = 392.55$, $p < 0.001$, $\eta_p^2 = 0.29$), with a smaller congruency effect for PC-10 (85.40 – 87.91 = 2.51) than PC-90 (70.41

– 92.70 = 22.29) items.

Within the LWPS, we observed a significant within-participant interaction between proportion switch and trial type ($F(1,955) = 47.07$, $p < 0.001$, $\eta_p^2 = 0.05$): the switch cost was reduced for PS-80 (MS first: 83.38 – 87.30 = 3.92; MR first: 86.46 – 90.62 = 4.16) compared to PS-20 (MS first: 85.61 – 90.41 = 4.8; MR first: 81.45 – 88.77 = 7.32) items, replicating the biased list/item effect. We also observed an interaction between block order (between-participants PS) and trial type ($F(1,955) = 46.11$, $p < 0.001$, $\eta_p^2 = 0.05$), with a smaller switch cost for PS-80 (83.38 – 87.30 = 3.92) than PS-20 (81.45 – 88.77 = 7.32) items.

Unbiased PC-50 and PS-50 Items

Reaction Time (ms). Within the LWPC, we observed a significant within-participant interaction between proportion congruency and congruency ($F(1,951) = 64.88$, $p < 0.001$, $\eta_p^2 = 0.06$): the congruency effect was reduced for the PC-25 (MI first: 696 – 644 = 52; MC first: 696 – 652 = 44) compared to PC-75 (MI first: 694 – 642 = 52; MC first: 717 – 651 = 66) contexts, replicating the unbiased LWPC effect. We also observed an interaction between block order (between-participants PC) and congruency ($F(1,955) = 28.19$, $p < 0.001$, $\eta_p^2 = 0.03$), with a smaller congruency effect for the PC-25 (696 – 644 = 52) than PC-75 (717 – 651 = 66) context.

Within the LWPS, we observed a significant within-participant interaction between proportion switch and trial type ($F(1,950) = 106.88$, $p < 0.001$, $\eta_p^2 = 0.10$): the switch cost was reduced for the PS-70 (MS first: 757 – 746 = 11; MR first: 734 – 717 = 17) compared to PS-30 (MS first: 736 – 709 = 27; MR first: 760 – 728 = 32) context, replicating the unbiased LWPS effect. We also observed an interaction between block order (between-participants PS) and trial type ($F(1,955) = 70.52$, $p < 0.001$, $\eta_p^2 = 0.07$), with a smaller switch cost for the PS-70 (757 – 746 = 11) than PS-30 (760 – 728 = 32) context.

Accuracy (%). Within the LWPC, we observed a significant within-participant interaction between proportion congruency and congruency ($F(1,955) = 38.53$, $p < 0.001$, $\eta_p^2 = 0.04$): the congruency effect was reduced for the PC-25 (MI first: 77.59 – 87.79 = 10.2; MC first: 83.89 – 89.33 = 5.44) compared to PC-75 (MI first: 82.46 – 89.80 = 7.34; MC first: 74.27 – 87.33 = 13.06) contexts, replicating the unbiased LWPC. We also observed an interaction between block order (between-participants PC) and congruency ($F(1,955) = 13.23$, $p < 0.001$, $\eta_p^2 = 0.01$), with a smaller congruency effect for the PC-25 (77.59 – 87.79 = 10.2) than PC-75 (74.27 – 87.33 = 13.06) context.

Within the LWPS, we observed a significant within-participant interaction between proportion switch and trial type ($F(1,955) = 15.28$, $p < 0.001$, $\eta_p^2 = 0.02$): the switch cost was reduced for the PS-70 (MS first: 81.75 – 86.33 = 4.58; MR first: 84.66 – 89.55 = 4.89) compared to PS-30 (MS first: 83.83 – 89.00 = 5.17; MR first: 80.03 – 87.07 = 7.04) context, replicating the unbiased LWPS. We also observed an interaction between block order (between-participants PS) and trial type ($F(1,955) = 20.74$, $p < 0.001$, $\eta_p^2 = 0.02$), with a smaller switch cost for the PS-70 (81.75 – 86.33 = 4.58) than PS-30 (80.03 – 87.07 = 7.04) context.

Explicit Awareness as a Learning Signal

The extent to which explicit awareness plays a role in control-learning remains debated (Abrahamse et al., 2016). Here, we assessed explicit awareness of the proportion context manipulations with a variety of questions. First, we simply asked participants, “When you were categorizing color-words/digits and letters, were some **blocks** (some of the **color-words/digits and letters**) harder to categorize than others?” With respect to blocks, participants self-reported more explicit awareness for the LWPC than the LWPS (LWPC, Yes: 607, No: 317, Don’t Know:

32, N/A: 1; LWPS, Yes: 389, No: 481, Don't Know: 85, N/A: 2). This pattern was also observed for specific color-words or digits and letters, albeit less pronounced (LWPC, Yes: 615, No: 315, Don't Know: 26, N/A: 1; LWPS, Yes: 522, No: 391, Don't Know: 42, N/A: 2). Whether participants were asked about the item-specific or block-level difficulty was counterbalanced, but participants were asked all the LWPC post-test questions before the LWPS questions, so it is possible this influenced subsequent responding. Nonetheless, when explicitly asked about the difficulty of task for either the item-specific or list-wide associations, participants largely endorsed explicit awareness of the task manipulation.

To further explore explicit awareness, we asked participants about the task statistics, stimuli, and strength of the proportion context manipulations. First, participants were asked to identify the incongruent color associated with each color-word (e.g., When the color-word RED was not printed in RED, it was printed in: …"). Participants performed better than chance (20%, $t(956) = 56.49$, $p < 0.001$, Cohen's $d = 1.83$), and their mean accuracy for identification differed between PC-90, PC-10, and PC-50 items (Supplementary Figure 1A; $F(1.96, 1874.28) = 59.16$, $p < 0.001$, $\eta_p^2 = 0.06$). Specifically, identification accuracy tracked with proportion of incongruent trials, with the highest accuracy for PC-10 items (82.7%), then PC-50 items (76.6%), and finally, PC-90 items (69.9%). Notably, of these item types, PC-90 is the most fluent in terms of perceptual processing, followed by PC-50, then PC-10. That participants identify PC-10, followed by PC-50, most accurately suggests that item-specific difficulty, not fluency, impacts explicit awareness here and results in feature-based selective attention such that participants can better identify frequent incongruent pairings.

Second, participants were asked how often each color-word was presented in its congruent color, with the scale anchored at 0 (Always presented in ANOTHER COLOR), 50

(Equally in RED and ANOTHER COLOR), and 100 (Always presented in RED). Participants underestimated the difficulty of PC-10 items (vs. 10%: $t(953) = 77.13$, $p < 0.001$, Cohen's $d = 2.50$, $M = 48.3\%$) and overestimated the difficulty of PC-90 items (vs. 90%: $t(953) = 54.62$, $p < 0.001$, Cohen's $d = 1.77$, $M = 61.5\%$) as well as PC-50 items (vs. 50%: $t(953) = 8.66$, $p < 0.001$, Cohen's $d = 0.28$, $M = 53.8\%$). Mean estimates were significantly different from each other (Supplementary Figure 1B; $F(1.77, 1689.94) = 317.37$, $p < 0.001$, $\eta_p^2 = 0.25$), despite the estimates clearly hovering near 50% across all items. The underestimation of difficulty for biased PC-10 items (38.3%) was larger than the overestimation of difficulty for biased PC-90 items (28.5%). Thus, although participants may be aware of stimulus pairings, they are not good at estimating the difference in difficulty that causes adjustments in control.

Third, as with other studies (Bejjani et al., 2018; Bejjani, Tan, et al., 2020; Bejjani & Egner, 2021), we asked participants to match the color-words to their respective difficulty levels (hard, easy, neutral). As shown in Supplementary Figure 1C, participants performed above chance (2/6 correct, $t(956) = 13.11$, $p < 0.001$, Cohen's $d = 0.42$, $M = 2.73$), largely driven by their ability to identify the biased color-words above chance (4/3, $t(956) = 15.92$, $p < 0.001$, Cohen's $d = 0.51$, $M = 2.08$) and not the unbiased color-words (2/3, $t(956) = 0.57$, $p = 0.571$, Cohen's $d = 0.02$, $M = 0.65$). We thus replicated results from previous studies, suggesting that while participants could identify that biased items are "easy" or "hard", they miscategorized unbiased items, not rating their difficulty level as "neutral". This suggests that increased explicit awareness of the difficulty associated with the biased items might act as a learning signal to generalize the attentional state to unbiased items.

Fourth, we asked participants to identify the temporal difficulty of the blocks within the LWPC and LWPS tasks (i.e., "There were 4/8 blocks in the categorization task involving color-

words/digits and letters. The first two/four blocks of the task differed from the last two/four blocks of the task in terms of the difficulty. Were the last two/four blocks you experienced **hard** or **easy**?" Participants could also select "Don't Know."). Supporting the assumption of increased awareness as a learning signal, we found that participants identified the temporal difficulty above chance (0.33) for both the LWPC ($t(956) = 10.67$, $p < 0.001$) and the LWPS ($t(956) = 7.35$, $p < 0.001$). As shown in Supplementary Figure 1D, they were better at identifying temporal difficulty for the LWPC ($M = 0.50$, 95% CI [0.47, 0.53]) than the LWPS ($M = 0.45$, [0.42, 0.48]; $t(956) = 2.44$, $p = 0.015$). Thus, although these post-test questions occurred at least 30 minutes after performing the LWPC, participants could still identify the most recent temporal difficulty.

Finally, participants were then told about the block manipulation and asked to estimate how often on hard/easy blocks, color-words were (not) printed in colors that matched their meaning or how much multitasking they thought they had to do. These judgments were anchored at 0 (Never matching, No multitasking), 50 (Balanced), 100 (Always matching/multitasking). Participants again underestimated the difficulty of LWPC hard blocks (vs. 25%, $t(954) = 41.75$, $p < 0.001$, $M = 49.10$ [47.97, 50.24]) and overestimated the difficulty of easy blocks (vs. 75%, $t(954) = 16.65$, $p < 0.001$, $M = 65.29$ [64.14, 66.43]), with a larger difference for hard (24.10%) than easy (9.71%) block estimation (Supplementary Figure 1E). They likewise overestimated the difficulty of LWPS easy blocks (vs. 30%, $t(954) = 24.90$, $p < 0.001$, $M = 47.05$ [45.70, 48.39]), but correctly estimated the hard blocks (vs. 70%, $t(954) = 1.96$, $p = 0.050$, $M = 68.78$ [67.56, 70.00]), potentially suggesting a difference between domain difficulty. Although participants can accurately report that a block was "harder" or "easier" than others, they again are not as accurate at estimating the degree of the block manipulation, with the exception of LWPS hard blocks.