Appendix A1: Supplementary Experiment

This auxiliary experiment uses the attention operating characteristic method (Sperling & Melchner, 1978; Sperling & Dosher, 1986), which tests the extent to which two tasks use the same attentional resources. We used a very similar approach to that described in detail by Alvarez and colleagues (2005). To briefly summarize the logic of this method, there are several possible scenarios for how two tasks might interact with one another. If two tasks use completely independent resources, performing the two tasks together should be no more difficult than performing each task individually (as would be predicted for tracking objects presented in separate visual hemifields; Alvarez & Cavanagh, 2005). Alternatively, if two tasks draw continuously from the same resource, performing two tasks together should result in a decrease in performance relative to performing the tasks individually, as performing better on the first task would require participants to remove resources from (and thus perform worse on) the second task (as would be predicted for tracking two items in the same visual hemifield; Alvarez & Cavanagh, 2005). The drop in performance when completing the two tasks simultaneously can be compared to the line of mutual exclusivity, which assumes a linear tradeoff between tasks. Performance that falls on or below this line of mutual exclusivity indicates that two tasks draw upon the same attentional capacity, while performance that falls above this line indicates that the tasks use at least partially independent resources (a pattern that has been found for simultaneously performing a search and tracking task, as well as for simultaneously performing a tracking and auditory task; Alvarez et al., 2005). Points that fall below the line of mutual exclusivity are expected when tasks have steep performance-resource functions (i.e., when you cannot perform well unless you give the task a lot of resources; see Alvarez et al., 2005), such that dividing resources between two task leads to poor performance on both.

In our primary training experiment, we tested for transfer 1) between the upper and lower visual fields, 2) between dot and pinwheel tracking, and 3) between both these factors at the same time. To test A) whether items in the upper vs. lower visual field draw upon the same capacity and B) whether the dot and pinwheel tasks draw upon the same capacity, in this supplementary analysis we performed three experiments testing whether 1) dots in the upper visual field and dots in the lower visual field draw continuously from the same capacity, 2) pinwheels in the upper visual field and pinwheels in the lower visual field draw continuously from the same capacity, and 3) whether dots and pinwheels draw continuously from the same capacity.

Each experiment followed the procedure used by Alvarez et al. (2005). In each of the three experiments, 80% speed thresholds were calculated separately for successfully tracking two targets in the upper and lower quadrant of the same hemifield (either 1) dots in both quadrants, 2) pinwheels in both quadrants, or 3) dots in one quadrant and pinwheels in the other). Within each quadrant, stimuli were identical to those used in the primary training experiment. Next, participants completed 5 conditions that were perceptually identical, but required differential attentional allocation: 1) track targets only in the upper visual field (top only), 2) track targets only in the lower visual field (bottom only), 3) track targets in both the upper and lower visual fields, but prioritize targets in the upper visual field (priority top), 4) track targets in both the upper and lower visual fields, but prioritize targets in the lower visual field (priority bottom), 5) track targets in both the upper and lower visual fields, and prioritize all targets equally (priority equal). For conditions 3-5 (where tracking occurred simultaneously in both the upper and lower visual field), we separately scored performance in the upper and lower visual fields relative to
tracking performance for each visual field individually (conditions 1-2), using the formula:

\[
\text{normalized score} = \frac{\text{(dual task accuracy – chance)}}{\text{(single task accuracy – chance)}} \times 100.
\]

For example, for both the dot and pinwheel task (where chance accuracy is 50% when identifying 2/4 targets), if tracking performance in the upper visual field for dots was found to be 90% when tracking only in the upper visual field, but only 70% when tracking in both visual fields with equal priority, the normalized score for dots in the upper visual field in the equal priority condition would be \[(70 - 50)/(90 - 50)] \times 100 = 50.

**Results**

Reported means refer to distance from the mutual exclusivity line. A negative mean refers to performance falling below the mutual exclusivity line, while a positive mean refers to performance falling above the mutual exclusivity line. Error bars represent within-subject SEM (Cousineau, 2005). Eight observers participated in each of the three experiments below.

**Experiment A1A: Dot Tracking in Both the Upper and Lower Visual Field (N = 8)**

- Performance was not significantly different than the mutual exclusivity line for either the priority top (\(M = -0.09, SD = 0.15, t(7) = 1.72, p = .13\)) or the priority bottom (\(M = 0.04, SD = 0.12, t(7) = 0.98, p = .36\)) conditions. However, performance for the priority equal condition was significantly below the mutual exclusivity line (\(M = -0.26, SD = 0.15, t(7) = 4.66, p = .002\)). All three points fell on or below the mutual exclusivity line, indicating that tracking dots in the upper visual field requires the same attentional capacity as tracking dots in the lower visual field.

**Experiment A1B: Pinwheel Tracking in Both the Upper and Lower Visual Field (N = 8)**
Figure A1B: Performance was not significantly different than the mutual exclusivity line for the priority bottom condition ($M = -.03$, $SD = .20$, $t(7) = 0.46$, $p = .66$). However, performance for both the priority top ($M = -.20$, $SD = .21$, $t(7) = 2.66$, $p = .03$) and the priority equal ($M = -.36$, $SD = .26$, $t(7) = 3.84$, $p = .006$) conditions was significantly below the mutual exclusivity line. Once again, all three points fell on or below the mutual exclusivity line, indicating that tracking pinwheels in the upper visual field requires the same attentional capacity as tracking pinwheels in the lower visual field. Furthermore, a mixed factors ANOVA predicting distance from the line of mutual exclusivity by condition (priority top, priority equal, priority bottom) and experiment group (dot task, pinwheel task) revealed an insignificant main effect of group ($F(1, 14) = 1.69$, $p = .22$) and an insignificant group x condition interaction ($F(2, 28) = .057$, $p = .94$), indicating similar mutual exclusivity between dots in the upper vs. lower visual field and pinwheels in the upper vs. lower visual field.

Experiment A1C: Simultaneous Dot and Pinwheel Tracking (N = 8)

This experiment had the same setup as the previous two, except participants always had dots in one visual field and pinwheels in the other (counterbalanced across participants). Therefore, we refer to the conditions as 1) priority dot, 2) priority equal, and 3) priority pinwheel.
**Figure A1C:** Although performance was numerically below the mutual exclusivity line for all three conditions, the priority pinwheel ($M = -0.03$, $SD = 0.20$, $t(7) = 0.30$, $p = 0.77$), priority dot ($M = -0.10$, $SD = 0.14$, $t(7) = 2.04$, $p = 0.08$), and the priority equal ($M = -0.16$, $SD = 0.21$, $t(7) = 2.20$, $p = 0.06$) failed to deviate significantly from the mutual exclusivity line. However, again all three points fell on or below the mutual exclusivity line, suggesting that tracking pinwheels requires the same attentional capacity as tracking dots. Furthermore, a one-way ANOVA comparing the distance of the priority equal condition from the line of mutual exclusivity across the three experiments found no reliable differences ($F(2,21) = 1.67$, $p = 0.21$), suggesting that tracking dots versus pinwheels in two different locations is as mutually exclusive as tracking either dots or pinwheels in two different locations.

References:


Appendix A2: Staircase procedures

A: General Procedures:
- Speeds for each motion type/location task combination (e.g., dots in the upper left quadrant) were determined using two independent, interleaved staircases. For each staircase, each correct trial (correctly identifying both targets) resulted in the staircase speed increasing by the current step size. Each time a staircase had two consecutive incorrect trials occur, the staircase speed decreased by 3X the current step size. A reversal trial occurred when either 1) a correct trial occurred after any two consecutive incorrect trials or 2) two incorrect trials occurred following any correct trial. The mean of each staircase’s final four reversals was designated at the overall staircase speed. The mean of the two overall staircase speeds was designated as the threshold speed for each motion type/location combination.

B: Assessment Day Staircase Parameters
- One staircase (SC1) always began at half the speed of the second staircase (SC2). Each staircase had a step size of 3X the final step size until 4 reversals occurred. The step size was then set at 2X the final step size until two more reversals occurred (6 total). After 6 reversals, each staircase was set to its final step size until 10 reversals were reached. Staircases for all task types were randomly interleaved, and selected based on a weighted probability of the number of remaining reversals until completion.
  - Dots
    - Starting Speeds: SC1 = 8.923 deg/s, SC2 = 17.846 deg/s
    - Final Step Size = 0.892 deg/s
    - Minimum Speed = 1.78 deg/s
  - Pinwheels
    - Starting Speeds: SC1 = 150 deg/s, SC2 = 300 deg/s
    - Final Step Size = 10 deg/s
    - Minimum Speed = 50 deg/s
- Example Staircase (assessment session, pinwheel top-right quadrant. Final four reversals for each staircases are marked):
C: Training Day Staircase Parameters
- Two independent randomly interleaved staircases were also used for the training sessions. Staircases ran for 55 minutes regardless of number of reversals. SC1 began a designated amount below the previous day’s threshold speed, while SC2 began a designated amount above (see below). For the first training session (day 3), the baseline assessment speed for the training task (from day 2) was used as the previous day’s threshold speed. For the rest of the training sessions (days 4-8), the mean of each staircase’s final four reversals was used to calculate threshold speed in the same way as in the assessment sessions (described above). The step size was constant throughout, and the same as the final step size used in the assessment sessions.
  - Dots
    - Starting Speeds: SC1 = previous day’s threshold – 4.46 deg/s, SC2 = previous day’s threshold + 4.46 deg/s
    - Step Size = 0.892 deg/s
    - Minimum Speed = 1.78 deg/s
  - Pinwheels
    - Starting Speeds: previous day’s threshold – 75 deg/s, SC2 = previous day’s threshold + 75 deg/s
    - Final Step Size = 10 deg/s
    - Minimum Speed = 50 deg/s
- Example Staircase (training session, pinwheel top-right quadrant. Final four reversals for each staircases are marked):
D: Practice Day Staircase Parameters

- On practice days, staircases were run for 30 minutes, regardless of number of reversals. Procedure for practice day was similar to the assessment day, except with different starting speeds and step sizes. The step size also remained constant throughout the practice session.

  o Dots
    - Starting Speeds: SC1 = 5.354 deg/s, SC2 = 10.707 deg/s
    - Step Size = 0.446 deg/s
    - Minimum Speed = 1.78 deg/s

  o Pinwheels
    - Starting Speeds: SC1 = 100 deg/s, SC2 = 200 deg/s
    - Step Size = 5 deg/s
    - Minimum Speed = 50 deg/s
Appendix A3: Results Before Subject Replacement

Figure A3: Training results if subject replacement had not occurred. The Trained condition still showed significantly greater improvement than all other conditions (including control), while all transfer conditions (New Motion, New Location, and Both New) all failed to differ significantly from control. Improvement for all conditions was still significantly greater than 0. The only change in significance was that the New Location condition was not significantly different than control before subject replacement occurred, but was significantly greater in our main report. The New Location was still significantly greater than the Both New condition (when aggregated across dot and pinwheel training), however, consistent with our claim that partially shared features between training and outcome measures may allow limited generalization of training benefits.
**Appendix A4:** Condition Means and Standard Deviations

**Table A4A:**

*Mean and Standard Deviation of Percent Improvement by Training Condition*

<table>
<thead>
<tr>
<th></th>
<th>Dot Training</th>
<th>Pinwheel Training</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Trained</td>
<td>59.6</td>
<td>33.9</td>
</tr>
<tr>
<td>New Motion</td>
<td>26.7</td>
<td>33.9</td>
</tr>
<tr>
<td>New Location</td>
<td>34.4</td>
<td>22.8</td>
</tr>
<tr>
<td>Both New</td>
<td>18.7</td>
<td>29.5</td>
</tr>
</tbody>
</table>

**Table A4B:**

*Mean and Standard Deviation of Control Condition*

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>17.4</td>
<td>20.3</td>
</tr>
</tbody>
</table>
**Appendix A5**: Individual Subject Distributions of Learning and Transfer

**A**

![Graph A](image)

**B**

![Graph B](image)
Figure A5: Individual Subject Distributions of Learning and Transfer. Each figure plots individual subject improvement for the trained task versus improvement for one of the untrained transfer tasks (A: new motion, B: new location, C: both new). Within each plot, each data point represents a subject in either the dot training (black circles) or pinwheel training (red circles) group. Points that fall within the white region indicate greater gains for the trained condition than the untrained transfer condition, while points in the gray indicate the opposite. Points falling on the line between the white and gray regions indicate equal improvement for the training and untrained transfer condition.
Appendix A6: Mixed Effects Analysis Code

The original R file can be downloaded at:
https://scorsese.wjh.harvard.edu/turk/experiments/rws/Misc/Exp3_Mixed_Effects.R

```r
#load packages
if (!require("lme4")) install.packages("lme4"); require("lme4")
if (!require("lmerTest")) install.packages("lmerTest"); require("lmerTest")
if (!require("effects")) install.packages("effects"); require("effects") #for plotting

#import data
file_link = 'https://scorsese.wjh.harvard.edu/turk/experiments/rws/Misc/MOT_Training_Data.csv'
data_matrix = read.table(file_link, header = TRUE, sep = "", )

---data_matrix (and data_matrix_noControl) explanation
### id = subject number
### training_group = which training group the participant was part of: Exp1 (dot training), Exp2 (pw training), or Exp3 (control)
### did_training = During training period, whether participants did training: Yes (Exp 1/2) or No (Exp 3)
### left_right = visual field of task: left or right
### motion = motion type of task: Dot or PW
### top_bottom = visual field of task: top or bottom
### pre = baseline speed threshold (in degrees of visual angle/s for Dot, in degrees/s of rotation for PW)
### post = post-training speed threshold (in degrees of visual angle/s for Dot, in degrees/s of rotation for PW)
### improvement = 100*((post-pre)/pre) - outcome measure of interest
### condition = condition of task
### - trained = trained task
### - new motion = same location as training task, but different motion
### - new location = same motion as training task, but different location
### - both new = both location and motion are different from training task
### - control = there was no training task (control participants)

#create subject id factor levels
data_matrix$id = factor(data_matrix$id)

#sort factor levels for condition to be 1) Trained, 2) New Motion, 3) New Location, 4) Both New, 5) Control (for plotting)
data_matrix$condition = factor(data_matrix$condition, levels(data_matrix$condition)[c(5,4,3,1,2)])

#make a matrix of training (non-control) subjects only for first part of analysis
data_matrix_noControl = data_matrix[data_matrix$did_training == "Yes",]

#reference conditions for statistical tests (Exp 1/2 = trained, Exp 3 = control)
data_matrix_noControl$condition <- relevel(factor(data_matrix_noControl$condition), ref = "trained")
data_matrix$condition <- relevel(factor(data_matrix$condition), ref = "control")
```
#--------- Model Comparison to Test For Differences Between Exp 1 & Exp 2 
# starting with fully saturated model with all possible fixed effects
m0 <- lmer(improvement ~ condition * training_group * top_bottom * left_right + (1 | id), data = data_matrix_noControl)
summary(m0)
#use step function in lmerTest to perform backward elimination of non-significant factors
m0_reduced = step(m0)
m0_reduced
#---> simplified model is: improvement ~ condition + (1 | id)...other fixed effects and interactions do not provide a better fit

#Model 1: DV = improvement | Fixed Effect: condition | Random Intercept: id
m1 <- lmer(improvement ~ condition + (1 | id), data = data_matrix_noControl)
summary(m1)
m1.eff <- Effect("condition", m1)
plot(m1.eff)
#Comparison of fixed effects (relative to trained condition):
#                         Estimate    Std.Error   df      t value   Pr(>|t|)
#(Intercept)               57.918      4.577    89.490    12.654    < 2e-16
#  conditionnew motion    -32.059      5.184    93.000   -6.184    1.65e-08
#  conditionnew location  -23.182      5.184    93.000   -4.472    2.19e-05
#  conditionboth new      -40.306      5.184    93.000   -7.774    1.00e-11
#----> improvement for trained condition significantly greater (p < .05) than other three conditions (new motion, new location, both new)

#Model 2: DV = improvement | Fixed Effects: condition, training_group | Random Intercept: id
#---> seeing if adding training_group as a fixed effect results in a better model
m2 <- lmer(improvement ~ condition*training_group + (1 | id), data = data_matrix_noControl)
summary(m2)
m2.eff <- Effect(c('condition', 'training_group'), m2)
plot(m2.eff)
#Model Comparison: m1 vs m2
anova(m1, m2) #p = .9939
#-->Model 2 is not significantly better than Model 1 (Model 1 is actually better (although not significantly), with a lower AIC)
#-->therefore, no significant differences in improvement between experiments 1 and 2

#--------- Now Test For Differences Between Exp 1+2 vs Exp 3 <---------- ___
# starting with fully saturated model with all possible fixed effects
m3_sat <- lmer(improvement ~ condition * motion * top_bottom * left_right + (1 | id), data = data_matrix)
summary(m3_sat)
#use step function in lmerTest to perform backward elimination of non-significant factors
m3_reduced = step(m3_sat)
m3_reduced
# Model 3: DV = improvement | Fixed Effect: condition | Random Intercept: id
m3 <- lmer(improvement ~ condition + (1 | id), data = data_matrix)
summary(m3)
m3.eff <- Effect("condition", m3)
plot(m3.eff)

# Comparison of fixed effects (relative to control condition):
#                     Estimate    Std. Error      df     t value     Pr(>|t|)
# (Intercept)           17.3736     4.7999     46.0000     3.620     0.000732
# conditiontrained      40.5445     6.6900     75.0200     6.060     5.03e-08
# conditionnew motion    8.4851     6.6900     75.0200     1.268     0.208608
# conditionnew location 17.3623     6.6900     75.0200     2.595     0.011365
# conditionboth new      0.2387     6.6900     75.0200     0.036     0.971637

# Model 3b: DV = improvement | Fixed Effects: condition, motion, condition*motion | Random Intercept: id
m3b <- lmer(improvement ~ condition*motion + (1 | id), data = data_matrix)
summary(m3b)
m3b.eff <- Effect(c('condition', 'motion'), m3b)
plot(m3b.eff)

anova(m3, m3b)

# Model 3b is not significantly better than Model 3 (Model 3 is actually better (although not significantly), with a lower AIC)
# therefore, improvement did not vary across conditions differently between dot movement and pinwheel movement
Appendix A7: Dot Training and Pinwheel Training Learning Curves

A: Dot Training

B: Pinwheel Training

Figure A7: Speed Thresholds and Learning Curve for Dot Training (A) and Pinwheel Training (B). A different staircase procedure was used for the training days (days 3-8) than was used during the two assessment days (days 2 and 9). Error bars represent SEM. Note that the control condition in each of these graphs reflects the same data (as well as the same data presented in the learning curve in the manuscript), while the other four conditions represent data specific to dot training (A) or pinwheel training (B) subjects.
Appendix A8: Correlation of Dot and Pinwheel Baselines Speed Thresholds

Figure A8: Correlation of baseline dot and pinwheel speeds for participants across all three experiments (N = 48). Each participant received two baseline speeds for each motion type (one in the upper visual field and one in the lower visual field). For each participant, the two baseline speeds for each motion type were averaged, producing a single dot speed and pinwheel speed used to compute this correlation.