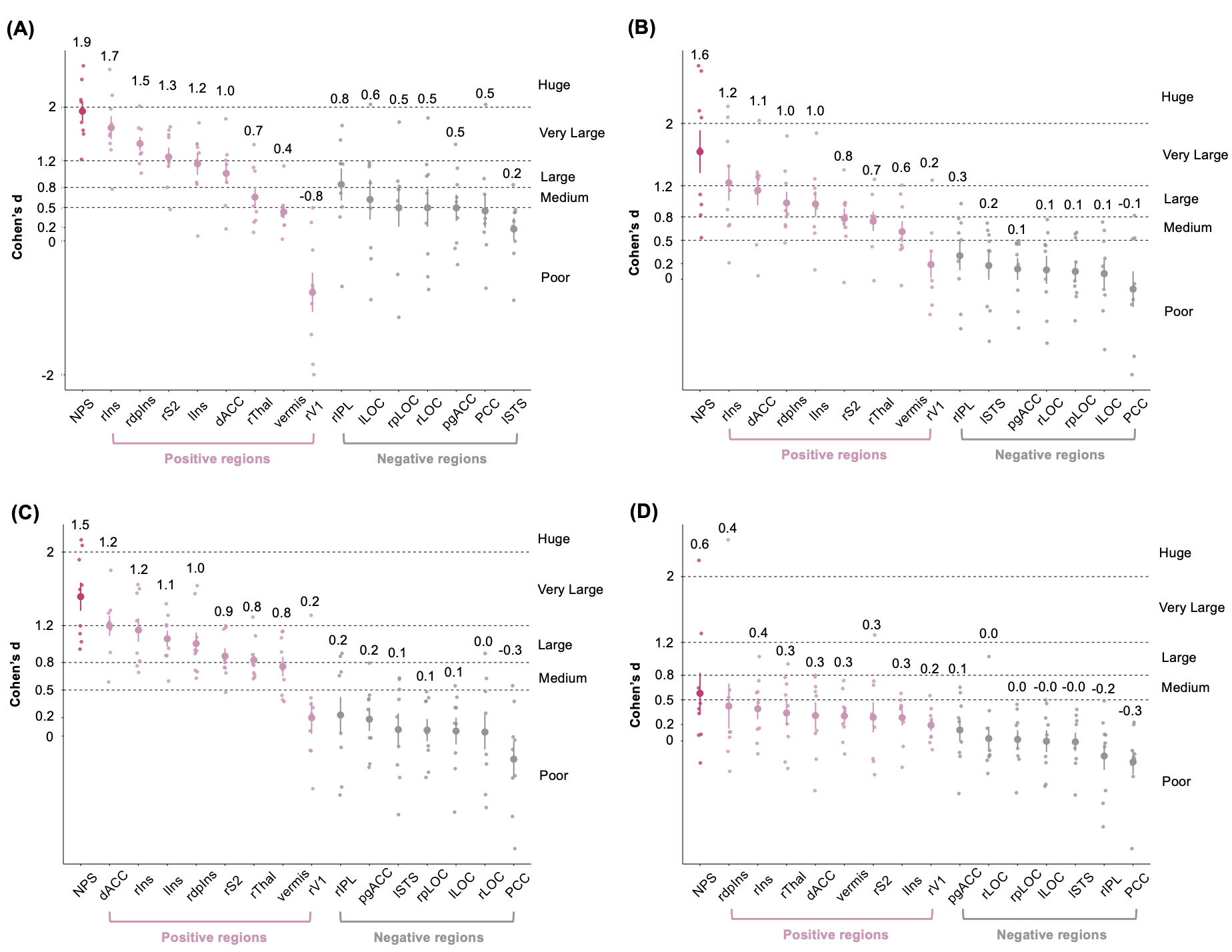
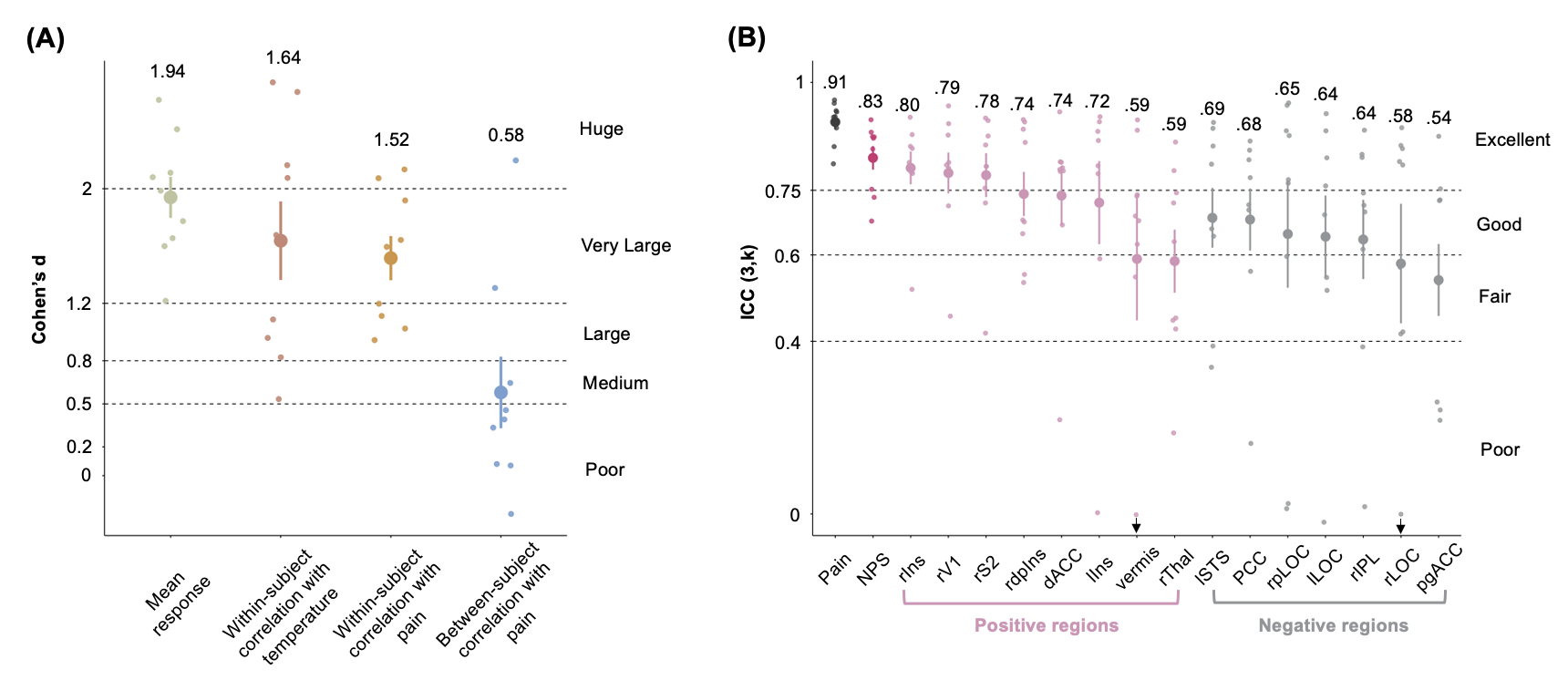


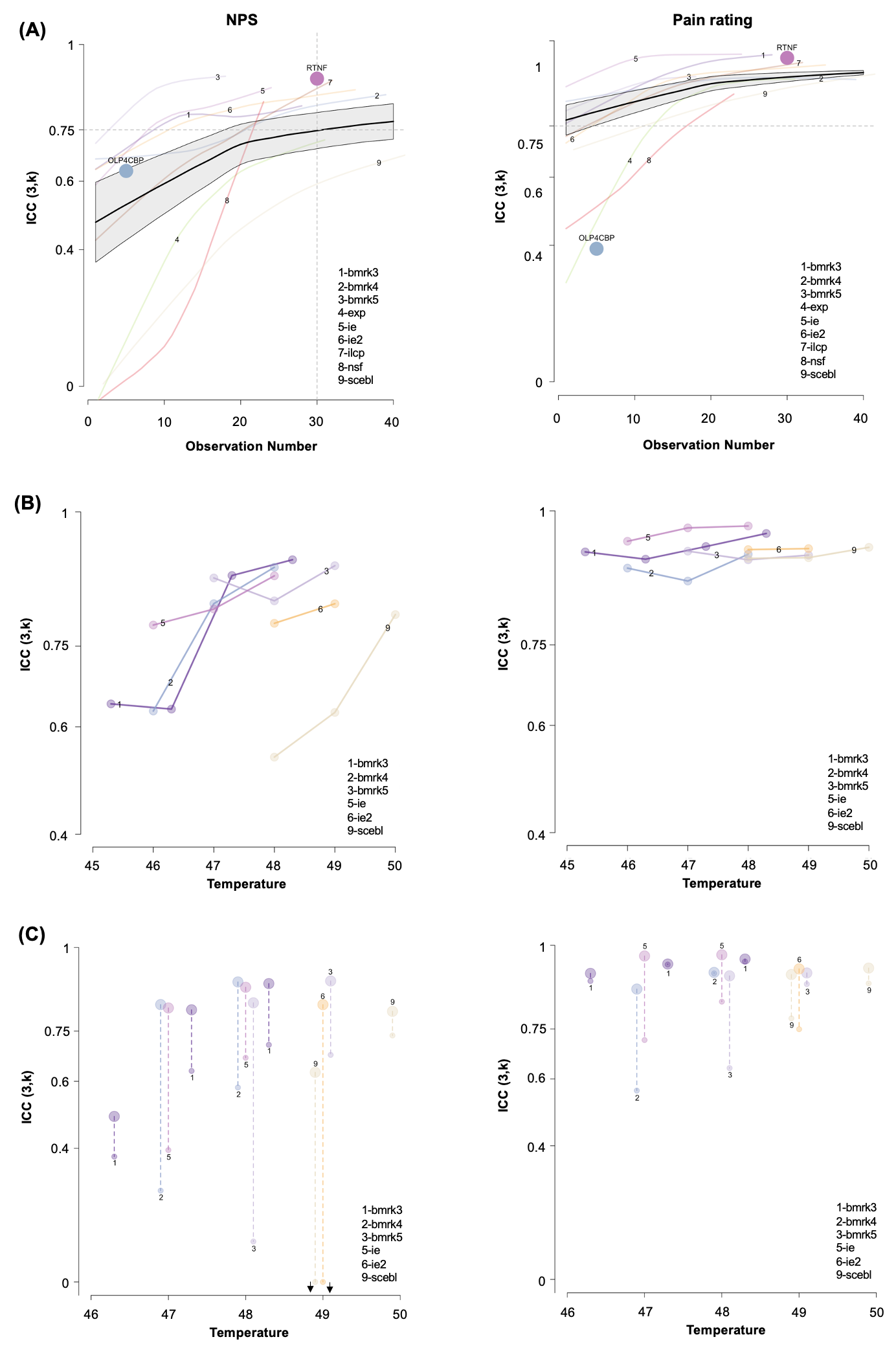
**Figure S1.** Four types of NPS effect. **(A)** Mean response of NPS. Each big dot represents the mean response of NPS in each study; the vertical bar represents the standard error; each small dot represents mean NPS response of one participant; and the violin plot represents the distribution of all participants in each study. To make the NPS response values comparable across different studies, the NPS response was rescaled by mean absolute deviation within each study. **(B)** Within-subject correlation between the NPS response and the temperature. Each big dot represents the mean beta value of the regression with the temperature as the independent variable and the NPS response as the dependent variable; the vertical bar represents the standard error; each small dot represents the beta value of one participant; and the violin plot represents the distribution of all participants in each study. **(C)** Within-subject correlation between the NPS response and the subjective pain ratings.Each big dot represents the mean beta value of the regression with the subjective pain ratings as the independent variable and the NPS response as the dependent variable; the vertical bar represents the standard error; each small dot represents the beta value of one participant; and the violin plot represents the distribution of all participants in each study. **(D)** Between-subject correlation between the NPS response and participants’ mean subjective pain ratings. Each dot represents one participant; and the line represents the linear regression relationship between the mean of the subjective pain ratings and the mean of the NPS response of each participant. \*\*\* p < 0.001; \* p<0.05.



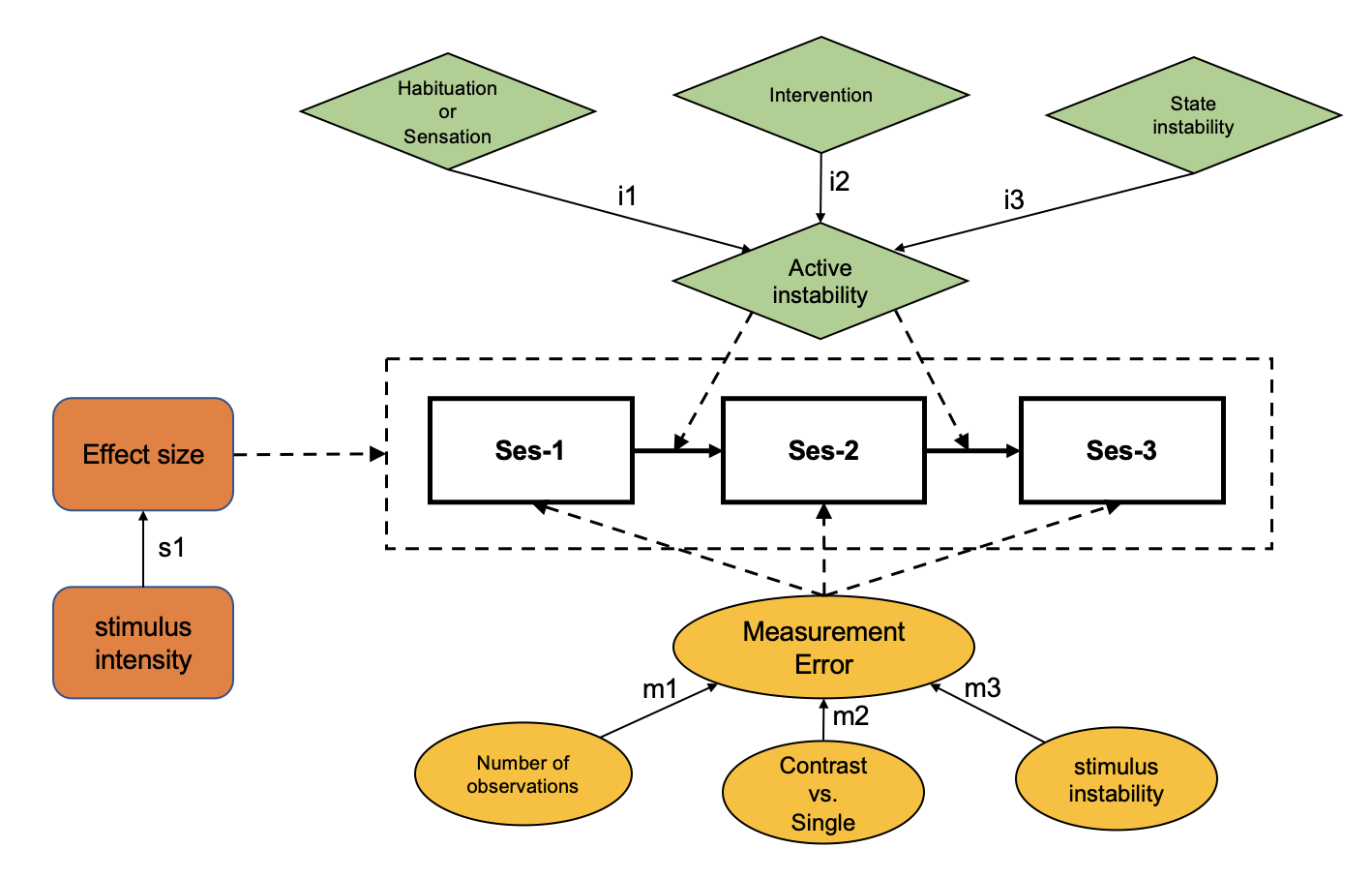
**Figure S2.** Four types of effect size of NPS and local areas. **(A)** Effect size of mean response of NPS and local areas. **(B)** Effect size of within-subject correlation with the temperature. **(C)** Effect size of within-subject correlation with the subjective pain ratings. **(D)** Effect size of between-subject correlation with the subjective pain ratings. Each big dot represents the mean effect size of the nine studies; the vertical bar represents the standard error; each small dot represents the effect size for one study. Ins denotes Insula, V1 primary visual area, S2 secondary somatosensory cortex, ACC anterior cingulate cortex, Thal thalamus, STS superior temporal sulcus, PCC posterior cingulate cortex, LOC lateral occipital complex, and IPL inferior parietal lobule. Direction is indicated with preceding lowercase letters as follows: r denotes right, l left, d dorsal, p posterior, pg perigenual.



**Figure 1.** Effect size and reliability of NPS. **(A)** Four types of effect size which are corresponding to the four NPS effect tests in Figure S1. **(B)** Reliability of subjective pain ratings, NPS pattern response and the pattern of local regions of NPS with positive or negative activations. Ins denotes Insula, V1 primary visual area, S2 secondary somatosensory cortex, ACC anterior cingulate cortex, Thal thalamus, STS superior temporal sulcus, PCC posterior cingulate cortex, LOC lateral occipital complex, and IPL inferior parietal lobule. Direction is indicated with preceding lowercase letters as follows: r denotes right, l left, d dorsal, p posterior, pg perigenual. The downward-pointing arrows indicate ICC < 0.



**Figure 2.** Factors that influence the reliability of NPS response (left column) and subjective pain ratings (right column). The small numbers from 1 to 9 correspond to the 9 single-trial study datasets. **(A)** Influence of the observation number and time interval between sessions. The ICC values are calculated based on different observation numbers (e.g., trials). Each colorful line shows the nonlinear relationship, which are fitted using *loess*, between the observation number and the ICC values in one of the nine single-trial datasets. If there are less than 10 participants under an observation number, it is excluded from the ICC calculation. The black line shows the average of the colorful lines corresponding to the datasets 1 to 8, which are weighted by the square root of the participants’ number. The grey shadow shows the standard error, which is also weighted by the square root of the participants’ number. On average, to achieve an excellent reliability, at least 30 trials are required to calculate the NPS response. **(B)** Influence of the temperature of the heat stimuli. If there are less than 13 participants under a temperature, it is excluded from the ICC calculation. With this criterion, the dataset 4, 7 and 8 are left with no ICC values in the plot. **(C)** Influence of the single condition vs. contrast. Larger dots represent the ICC values calculated by the NPS response (left panel) or subjective pain ratings (right panel) for each temperature; and the smaller dots represents the ICC values calculated by the NPS response or subjective pain ratings of each temperature minus the response to the lowest temperature in each dataset. The length of the dashed line represents the difference between the ICC values for high-temperature conditions vs. contrasts across temperatures. The downward-pointing arrow indicates ICC < 0. The contrast is less reliable than the NPS response or rating at a single temperature in virtually every case.



**Figure 3.** Summary of factors that influence reliability.