Online Supplemental Materials

CONSORT flow diagram

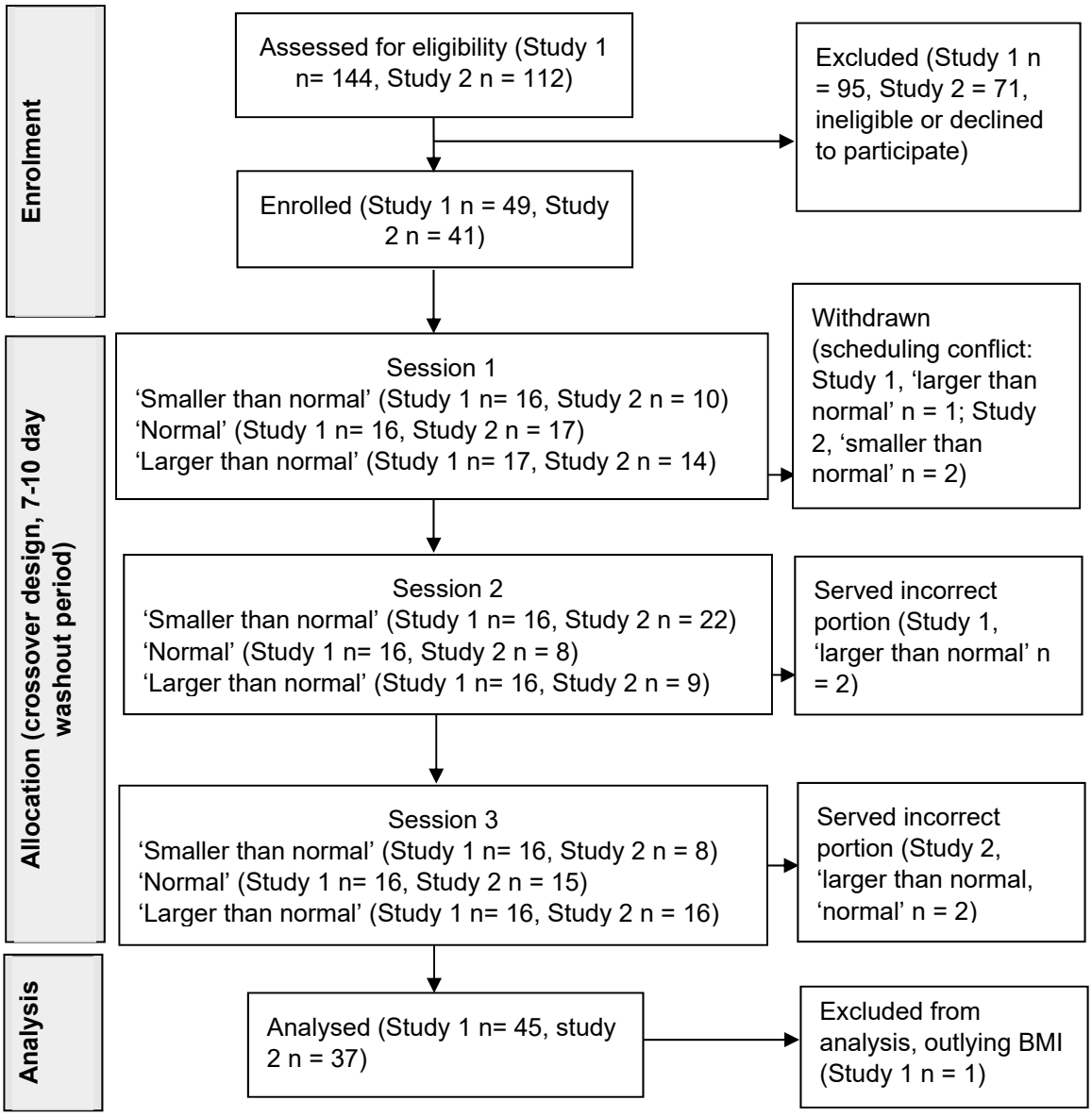

Figure S1. CONSORT Flow Diagram. *Note*. Values *n* in parentheses denote participants allocated to each condition in each session.

Table S1

Nutritional value of study foods (per 100g)

	Pasta with tomato sauce	Chicken curry	Long grain rice (cooked)	Caramel Shortbread	Flapjack bites
Energy	456 kJ	377 kJ	699 kJ	2092 kJ	1916 kJ
Fat	0.7g	3.7g	2.4g	28.0g	24.0g
(Saturated)	0.2g	0.4g	0.3g	15.2g	10.1g
Carbohydrate	22.0g	1.8g	31.7g	55.6g	54.5g
(Sugars)	1.6g	0.5g	0.2g	36.6g	30.3g
(Fibre)	0.5g	2.5g	1.9g	2.5g	4.2g
Protein	3.6g	11.0g	3.7g	5.1g	6.0g
Salt	0.3g	0.6g	0.1g	0.5g	0.4g

Filler measures

To bolster the cover story in Study 1, participants completed a 'word categorisation task' before and after consuming their lunchtime meal in each session. The task presented a series of words one by one in a random order, which participants categorised using keys marked 'left' and 'right' on the keyboard (see Table S2). Participants completed a different word task in each session, and task assignment to meal session was counterbalanced.

In Study 2, participants were asked to complete additional mood-related questions after finishing their meal in each session in keeping with the cover story about 'mood stability'. Participants indicated on 7-point Likert scales the extent to which they agreed with four statements about their mood during the session (e.g., "my mood changed after eating lunch today", "I feel energised after eating lunch today"), and four statements about their

mood over the past 7 days (e.g., “my mood has been mostly positive”, “my mood has been mostly stable”).

Table S2.

Word categorisation task stimuli (Study 1).

Task	**Word lists**	
Word / non-word	ripple	frine
	flood	meagle
“indicate whether each letter	youth	wie
string is a word or a non-word”	towel	glear
	carpet	ligget
	paper	sinth
Manmade / natural	desk	eagle
	bicycle	tulip
“indicate whether each word	shampoo	pine
refers to a manmade or natural	key	weasel
object”	paint	granite
	cement	mountain
Noun / verb	blackboard	compute
	doormat	imagine
“indicate whether each word	chair	write
refers to a noun (i.e., an object,	sky	open
or ‘thing’) or a verb (i.e.,	window	wash
describes an action)”	brick	fabricate

Additional self-report measures

The following measures are part of a standard questionnaire battery that we routinely collect in laboratory studies on eating. We did not intend to use these measures in the main analyses in relation to the hypotheses tested in Studies 1 and 2.

Awareness of portion size influence

Participants were asked whether they thought that the portion size of food served during each session influenced how much they ate, and responded a 7-point Likert scale (from 1, strongly disagree, to 7, strongly agree).

Awareness of portion size manipulation

Participants were asked to indicate whether they thought the amount of food served in each session differed (yes/no). To distract from the focus on portion size, participants completed several filler questions about their study experience (six yes/no questions in Study 1, four in Study 2, e.g., "I reported my mood before and after the meal each week", "the word categorisation computer tasks presented different words each week").

Income and education

Participants were asked to indicate their highest level of education *(did not complete high school, high school, some university, Bachelor's degree, Master's degree, Doctoral or professional degree*), and to enter their total annual household income (free text entry).

Food liking

Two questions were administered to assess general and study-specific liking of the foods served (Study 1: pasta with tomato sauce, Study 2: chicken curry with rice, millionaire shortbread, flapjack). Participants indicated on 7-point Likert scales how much they liked

each food in general and specifically when served during the study, ranging from 1 (*not at all)*, to 7 (*like it a lot*).

Restrained, external, and emotional eating

Participants completed the commonly used Dutch Eating Behaviour Questionnaire (DEBQ; Van Strien et al., 1986), which consists of 33 items in three subscales measuring restrained, external, and emotional eating tendencies. The scales have been demonstrated to have good internal reliability (Van Strien et al., 1986).

Dieting status

A single item was administered to assess whether participants were currently dieting (yes/no).

Plate clearing tendency

A five item measure with a 5-point Likert response format (1 [*strongly disagree*], to 5 [*strongly agree*]) was used to assess participants' tendencies to clear their plate when eating (e.g., " I always clear my plate when eating"; Robinson et al., 2015). The scale has been demonstrated to have good internal reliability (Robinson et al., 2015).

Self-control

The brief self-control scale was used to assess trait self-control (De Ridder et al., 2011). Participants responded to 13-items on 5-point Likert scales ranging from 1 (*strongly disagree*), to 5 (*strongly agree*) (e.g., "I am good at resisting temptation"). The scale has good internal reliability (De Ridder et al., 2011).

Attitudes against food waste

Participants completed a five item measure with a 7-point Likert response format (1 [*strongly disagree*], to 7 [*strongly agree*]) assessing attitudes against wasting food (e.g., "Even if I felt full, I would rather finish what is on my plate than see it go to waste".

Secondary analysis: hunger and fullness

From pre- post meal in each study, hunger significantly decreased, $F(1, 44) = 397.98$, $p < .001$, partial $\eta^2 = .90$ (Study 1), $F(1, 36) = 417.30$, $p < .001$, partial $\eta^2 = .92$ (Study 2), and fullness significantly increased, $F(1, 44) = 447.57$, $p < .001$, partial $\eta^2 = .91$ (Study 1), $F(1, 36) = 533.88$, $p < .001$, partial $\eta^2 = .94$ (Study 2). There was no significant interaction between condition and time (pre-post) on hunger, $F(2, 72) = 0.20$, $p = .82$, partial $\eta^2 = .005$, or fullness, $F(2, 72) = 2.65$, $p = .08$, partial $\eta^2 = .07$ in Study 2, but in Study 1 the interactions between condition and time (pre-post) on hunger, $F(1.76, 77.58) = 5.90$, $p = .01$, partial $\eta^2 = .12$, and fullness, $F(2, 88) = 5.49$, $p = .01$, partial $\eta^2 = .11$ were significant (Table 1, main manuscript). Separate repeated-measures ANOVAs to follow up the interaction terms in Study 1 also revealed significant differences in pre-meal hunger, $F(2, 88) = 6.04$, $p = .003$, partial $\eta^2 = .12$, and fullness, $F(2, 88) = 3.83$, $p = .03$, partial $\eta^2 = 08$ between conditions, but not for post-meal hunger, $F(1.69, 74.53) = 0.52$, $p = .57$, partial $\eta^2 = .01$, or fullness $F(2, 88) = 2.54$, $p = .09$, partial $\eta^2 = .06$. Mean pre- and post-meal hunger and fullness values, and results of pairwise comparisons for Study 1 values are reported in Table 1 (main manuscript).

As there were significant differences in pre-meal hunger and fullness between conditions in Study 1, three separate linear-mixed models were conducted to assess the effect of condition on intake variables (served portion, additional, and total meal energy intake), controlling for pre-meal hunger and fullness. In line with results of the main analysis, there was a significant effect of condition on energy intake from the served portion after controlling for pre-meal hunger and fullness, $F(2, 49.81) = 787.75$, $p < .001$. With each successive reduction in portion size, there was a significant reduction in intake from the served portion, *m* difference 'large-normal' to 'small-normal' reduction = 268.1 kJ, SE = 23.9, $p < .001$, $d = 1.68$, *m* difference 'small-normal' to 'smaller than normal' reduction =

302.5 kJ, SE = 10.1, p <.001, d = 4.45. There was also a significant effect of condition on additional energy intake after controlling for pre-meal hunger and fullness, $F(2, 99.96)$ = 7.22, p = .001. Also consistent with the main analysis, additional intake was significantly greater in the 'smaller than normal' than the 'small-normal' condition, *m* difference = 204.5 kJ, SE = 91.4, p = .03, d = 0.33, but did not significantly differ between the 'small-normal' and 'large-normal' conditions, *m* difference = 105.1 kJ, SE = 73.4, p = .16, d = 0.21. Controlling for pre-meal hunger and fullness, condition had a significant effect on total meal energy intake, $F(2, 97.58) = 4.78$, p = .01, and consistent with predictions (but contrary to the results of primary analyses), total meal intake did not significantly differ between the 'smaller than normal' and 'small-normal' conditions, *m* difference = 97.2 kJ, SE = 92.4, p = .30, d = 0.16, but was marginally significantly reduced in the 'small-normal' relative to the 'large-normal' condition, *m* difference = 156.9 kJ, SE = 80.3, p = .054, d = 0.29.

Analysis of order effects

To explore potential order effects we examined whether including a 6-level between-subjects factor of 'portion size sequence' affected the results of the primary analyses. In Study 1, the interaction between portion size condition and portion size sequence did not predict energy intake from the served portion, $F(10, 78) = 1.43$, p = .18, partial η^2 = .16, additional energy intake, $F(10, 78) = 1.29$, p = .25, partial η^2 = .14, or total energy intake, $F(10, 78) = 1.68$, p = .10, partial η^2 = .18. Statistically controlling for portion size sequence did not change the pattern of significance of primary analyses of additional and total energy intake, suggesting that the results of primary analyses were not largely dependent on the sequence in which the portions were served

In Study 2, the interaction between portion size condition and portion size sequence did not significantly predict energy intake from the served portion, $F(10, 62) = 1.02$, p = .44,

partial $\eta^2 = .14$, additional dessert intake, $F(10, 62) = 0.85$, $p = .58$, partial $\eta^2 = .12$, or total energy intake, $F(10, 62) = 0.77$, $p = .66$, partial $\eta^2 = .11$. The pattern of results in primary analyses remained the same when condition sequence was controlled for, except for additional energy intake, where unlike in primary analyses the smaller than normal and small-normal conditions did not significantly differ, *m* difference = 135.0 kJ, SE = 80.7, $p = .11$.

Participants' usual portion sizes

Participants' self-reported usual portion size of pasta with tomato sauce was largely consistent with the categorisation of portion sizes in Study 1 as it fell between the 'small-normal' and 'large-normal' portions: median = 100% of manufacturer's recommendation (IQR: 80 – 130%), portion sizes served: 60%, 90%, 120%. In Study 2 the median usual portion size of chicken curry with rice was closer to the 'smaller than normal' than the 'normal' portion sizes served during the study, median = 80% of manufacturer's recommendation (IQR: 70 - 100%), portion sizes served: 70%, 100%, 130%.