



*J. R. Statist. Soc. A* (2017)  
**180**, Part 1, pp. 225–246

# Unpacking the determinants of life satisfaction: a survey experiment

Viola Angelini,

*University of Groningen, and Netspar, Tilburg, The Netherlands*

Marco Bertoni

*University of Padova, Italy*

and Luca Corazzini

*University of Messina, and Bocconi University, Milan, Italy*

[Received August 2014. Final revision January 2016]

**Summary.** We present results of a survey experiment aimed at assessing context effects on reporting life satisfaction, exerted by raising awareness of fundamental life domains before eliciting overall life satisfaction, through questionnaire manipulations. Psychologists refer to similar context effects, generated by providing more details about the object of a subsequent evaluation, as ‘unpacking effects’. The longitudinal structure of our experimental design allows us to assess the effects of the questionnaire manipulation both between and within subject. In our sample of university students, asking subjects to report satisfaction with life domains before reporting overall satisfaction with life generates a robust unpacking effect, as it shifts upwards the subsequent mean overall life satisfaction evaluations. In addition, raising awareness about life domains significantly increases reliability and validity of self-reported life satisfaction, by reducing the dispersion of responses and increasing the association between life satisfaction and life domain evaluations. We also detect heterogeneous effects across subgroups of our sample—such as people with children or in bad health—and discuss implications of these findings for research on life satisfaction.

**Keywords:** Life satisfaction; Survey experiments; Unpacking effects

‘It isn’t what you have or who you are or where you are or what you are doing that makes you happy or unhappy. It is what you think about it.’  
Dale Carnegie (1936) *How to Win Friends and Influence People*

## 1. Introduction

The inability of objective economic indicators (such as *per capita* gross domestic product, real wages and financial wealth) to account fully for important aspects of citizens’ lives has driven social scientists to develop novel measures of subjective wellbeing (for a general discussion on the economic relevance of subjective wellbeing measures, see Diener (1984)). Among many approaches, those relying on data from representative general surveys are the most promising and commonly used (see Frey and Stutzer (2002a, b), van Praag and Ferrer-i-Carbonell (2004), Ferrer-i-Carbonell and Frijters (2004), Bruni and Porta (2005), Di Tella and MacCulloch (2006), Clark *et al.* (2008) and Dolan *et al.* (2008) for surveys). In these surveys, subjects are presented

*Address for correspondence:* Marco Bertoni, Department of Economics and Management, University of Padova, via del Santo 33, 35123 Padova, Italy.  
E-mail: marco.bertoni@unipd.it

with a large number of questions concerning their socio-economic, demographic and health conditions. In addition, they are asked to self-report their life satisfaction on an ordered scale going from very dissatisfied to very satisfied.

Standard assumptions of survey studies posit that respondents consciously and correctly report subjective information (see Schwarz and Strack (1999)), and that this information is comparable interpersonally or over time for the same person (see Angelini *et al.* (2014)). Of course, such assumptions are not innocuous. In particular, evaluating the level of satisfaction with one's own life is a complex task that requires sophisticated cognitive processes to make sense of the question asked, to build up a mental image of one's life and to formulate and report adequate judgements (Schwarz and Strack, 1991). In this perspective, it is reasonable to expect that format elements such as the framing, number and order of questions, as well as the information that is inferred from preceding tasks, exert strong psychological effects on subjects' responses and influence their informative content, with implications for economic research that either studies the determinants of satisfaction with life or considers subjective measures as determinants of relevant economic outcomes (see Bertrand and Mullainathan (2001) and Benjamin *et al.* (2012)). The literature, reviewed in the next section, refers to these psychological survey artefacts as 'context effects'.

In this paper, we present results of a survey experiment aimed at assessing context effects in subjects' reported satisfaction with life, exerted by raising awareness of fundamental life domains—income, family, job, friends, sentimental relationships and health—through questionnaire manipulations. Psychologists refer to context effects that are generated by providing more details about the object of a subsequent evaluation as 'unpacking effects' (see Van Boven and Epley (2003)). In our experiment we compare overall life satisfaction evaluations from a benchmark questionnaire with no reference to life domains with those reported in two different questionnaires containing explicit reference to life domains. In both variants, before reporting their life satisfaction, subjects are presented with a list of the six life domains. However, only in the second variant are subjects required to evaluate their satisfaction with each single life domain.

Context effects are likely to influence the distribution of life satisfaction in several ways and are also likely to affect its reliability—i.e. the ability of life satisfaction to produce stable and consistent results—and its validity—i.e. its ability to measure the concept that it is intended to measure (see Cronbach (1971) and Cozby (2001)). First, we expect the mean of the distribution of life satisfaction to change because the unpacking manipulation leads individuals to gather more information about the topic to be evaluated, and—according to the psychological theory behind the unpacking effect—this might induce subjects to report more extreme evaluations on the ordered scale that is used to express responses. Since most of our sample is composed of relatively young, middle-class, well-educated and healthy individuals, we expect unpacking effects to affect the mean of life satisfaction in our sample positively. Second, helping subjects to build a more accurate representation of their satisfaction with life could result in a lower dispersion of subjective evaluations, reducing the variance of the distribution of life satisfaction and increasing its reliability. Finally, the overall evaluation could better reflect a weighted aggregation of the life domains when their salience is enhanced and subjects are induced to think more deeply about all the determinants of life satisfaction, thus increasing the validity of this measure. To help us to formalize these predictions, we also develop a simple quantitative model of weighted domain satisfaction.

Although the psychological literature is rich in contributions that analyse how the structure of the questionnaire affects subjects' responses, as far as we know there is no study that seeks to measure the empirical relevance of each of the context effects that were mentioned above in

reporting life satisfaction. Moreover, a novel feature of our design is that we study how context effects change when the intensity of the awareness manipulation is gradually increased across versions of questionnaire, starting from one with no reference to the life domains, moving to one that involves a weak manipulation in which subjects are simply presented with the list of life domains and, in turn, considering one that introduces a stronger manipulation, requiring subjects to evaluate their satisfaction with each domain before answering the general question. Finally, no previous contribution exploits experimental variation in treatment assignment both between subjects and within subjects over time.

Although simply presenting subjects with a list of life domains does not alter their responses, we detect significant context effects on both the level and the precision of the life satisfaction evaluations in the questionnaire with the strong awareness manipulation. Raising awareness about life domains reduces the variance of life satisfaction and increases the average of the evaluations reported, that we also find to be more correlated with satisfaction with domains.

We also show that the unpacking effect is heterogeneous with respect to respondents' characteristics: for instance, it is stronger for those reporting being in bad health. A possible interpretation of this result is in terms of a 'focusing illusion' effect (see Schkade and Kahneman (1998)), whereby unpacking leads people to consider their life in general, without focusing on a single negative aspect, which would have otherwise been overweighted in the overall evaluation. This result is especially relevant for research on the determinants of life satisfaction, as it illustrates that context effects not only reduce precisions of the estimates, because of the higher variance, but also that in a subjective wellbeing regression coefficients related to the variables for which we detect heterogeneous effects could change depending on the version of the questionnaire used.

The paper unfolds as follows. In the next section, we review the relevant psychological and economic literature dealing with context effects. In Section 3, we present our experimental design and state testable predictions by developing a formal model of weighted domain satisfaction, that helps us to formalize the outcomes of the theories put forward by the psychological literature. In Section 4, we discuss our statistical approach and present the results on the level and variance of reported life satisfaction, as well as on the association between overall evaluations and satisfaction with the six domains. Finally, in Section 5, we discuss the relevance of our findings for research on life satisfaction and conclude.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

<http://wileyonlinelibrary.com/journal/rss-datasets>

## 2. Literature review

This paper builds principally on the literature on context effects in the elicitation of attitudes in surveys. Tourangeau and Rasinski (1988) and Schwarz (1999) have provided thorough reviews of the relevant literature, and Schwarz and Strack (1999) focused in particular on subjective wellbeing questions. In our case, context effects can be framed within models depicting the cognitive process that respondents rely on when asked about satisfaction with life. Schwarz and Strack (1991) highlighted how such questions require respondents to carry out an almost impossible task in a very limited amount of time: in a few seconds (Diener *et al.*, 2000) they ought to make sense of the question asked, retrieve the relevant information, make a judgement, report it in accordance with the alternatives provided by the researchers and, in some instances, adjust it to match criteria of social desirability. As a consequence, reported levels of life satisfaction

can be thought of as spot judgements, mainly based on information that is accessible at that point in time (Schwarz and Strack, 1999), as individuals truncate the information search process as soon as they have collected enough information to formulate a judgement (Schwarz, 1999). By affecting the accessibility and salience of the information that respondents use to build up a mental image of their life, prior items asked in the questionnaire may provide a framework to respond to later questions, generating context effects.

In line with this literature, we expect that information primed by listing or asking people to evaluate satisfaction with specific life domains could increase the reliability of the information about overall life satisfaction, as a more thorough information seeking process ought to be carried out by respondents. The nature of the information primed is also relevant to generate context effects, as not all of it is used to formulate subsequent judgements. Strack *et al.* (1985) showed that life satisfaction evaluations of respondents who are asked to recall three positive life events before answering the general questions are higher than those of respondents who are asked to recall three negative life events. On the basis of an experiment where subjects are asked about satisfaction with their dating lives and with life in general, Strack *et al.* (1988) showed that, when a specific and a general question are placed within a conversational context where the researcher shows interest in both domains separately, information competing with the domain elicited by the specific question might be disregarded when answering the general question, as it might be considered as redundant (Tourangeau *et al.* (1991) referred to this as a 'subtraction' effect, whereas Schwarz *et al.* (1991) interpreted it as a 'contrast' effect). In this sense, the two different manipulations that we carry out may be less or more salient in terms of information retrieval and may have differential effects on the informational content of the general question.

Information primed by previous questions may not affect life satisfaction evaluations only in the information retrieval phase. Schwarz and Strack (1999) highlighted how context information also affects the respondent's understanding of the meaning of the life satisfaction question: does the researcher mean life as it was, as it is now or as it will be? What aspects of life are of interest for the researcher? Similarly, McClendon and O'Brien (1988) argued that, by providing a clear frame of reference, contextual information should allow people to reduce measurement error due to 'guessing', and thus to increase the reliability of the judgement expressed.

Conti and Pudney (2011) analysed context effects in the reporting phase, related to labelling of the answer categories and with face-to-face interviews *versus* self-completion questionnaires. Exploiting exogenous variation in the labelling of the categories of a job satisfaction question across waves in the British Household Panel Survey, they found that women are less likely than men to tick a response option that is numerically but not textually labelled, because of different preferences towards verbal *versus* numerical communication across genders. Furthermore, they showed that oral interviews and the presence of children during interviews produce more positive satisfaction judgements (the 'let's put on a good show for the interviewer' and 'not in front of the children' effects), and that women report lower job satisfaction if the partner is present during the interview, to conform with social norms about gender roles (the 'don't show your partner how satisfied you are' effect).

As general life satisfaction evaluations are carried out on the spot, mood state effects might be present as well, and it has been shown that mood affects general questions more than questions that are related to specific life domains. For instance, Schwarz and Clore (1983) showed that satisfaction with life in general is lower for individuals who are asked about it on a rainy day, although the robustness of these findings is called into question by more recent research (Lucas and Lawless, 2013). Deaton (2012) and Deaton and Stone (2013) showed instead that asking political questions before life satisfaction leads people to report lower levels of satisfaction, and that different subgroups of the population are affected in different ways, thus changing wellbeing

rankings across groups. Diener *et al.* (2000) instead highlighted how the general ‘positivity’ of respondents might affect evaluations of life satisfaction, as these may reflect latent dispositional tendencies more than evaluative judgements. This point was raised in a different flavour also by Clark *et al.* (2005) and Angelini *et al.* (2014) who stressed that people might attach the same label to different concepts of wellbeing, hampering interpersonal comparability of subjective wellbeing evaluations because of differences in reporting styles.

From a different perspective, enlisting life domains that are relevant for satisfaction with life or asking respondents to evaluate their satisfaction with these domains before expressing a general evaluation can generate what Van Boven and Epley (2003) called ‘unpacking effects’. In their view, presenting more detailed descriptions of a given event may change the subjective perception that people hold of it and make it more extreme. In a set of experiments, they showed that people are less prone to give mild evaluations when they are presented with or asked to generate more detailed descriptions of a situation they are asked to evaluate (e.g. sludge burning operations causing ‘all kinds of respiratory diseases’ *versus* ‘asthma, lung cancer, throat cancer and all kinds of respiratory diseases’). According to Van Boven and Epley (2003), unpacking a description makes it easier for people to recall features of the object of the evaluation that they would not have considered otherwise, to generate a better mental image of the situation and to think more in detail about the event. All these features allow respondents to gain awareness and to come forward with more extreme evaluations, that they would have probably not dared to give without a thorough comprehension of the topic to be assessed. Our survey experiment is the first case of unpacking manipulations in a life satisfaction questionnaire, and we expect to find similar results to those described above.

Van Boven and Epley (2003) claimed that unpacking effects extend the support theory for probability judgements that was proposed by Tversky and Koehler (1994) to evaluative judgements. Support theory states that detailing events into their disjoint components increases the overall perceived probability that the event may occur (for example the perceived probability of death by meningitis is lower than the sum of the perceived probabilities of death by viral meningitis or bacterial meningitis). A related psychological distortion is the so-called ‘identifiable victim’ effect: Small and Loewenstein (2003) and Kogut and Ritov (2005) showed that individuals donate more to help an identifiable victim than a statistical victim, whereas Small and Loewenstein (2005) found greater anger towards identifiable than to general perpetrators. Finally, in economics, part–whole bias in contingent valuation is another similar phenomenon: when asked to evaluate them separately, people are willing to pay more for the separate components of a good than for the bundled solution. Bateman *et al.* (1997) experimentally proved that this principle holds for the case of a drink and a burger *versus* a fast food menu made of the two goods, whereas Bernasconi *et al.* (2009) showed that unpacking a single public good into two components increases private contributions to support its provision.

### 3. The survey experiment

#### 3.1. Experimental design

This study aims at assessing whether raising awareness of important life domains affects how subjects evaluate their satisfaction with life. In a nutshell, we manipulate the level of awareness by administering three versions of a baseline questionnaire on life satisfaction: one including no reference to the life domains (henceforth indicated as T1), one simply presenting the list of the domains (T2) and, finally, one that requires subjects to evaluate satisfaction with each domain explicitly (T3). In our experiment we refer to six life domains—income, family, job, friends, sentimental relationships and health—that the literature has identified as main determinants of life



not required to express any explicit evaluation about the domains. Then, subjects report their overall life satisfaction on a 10-point scale. Apart from this difference, the design of treatment T2 replicates that used in treatment T1. In particular, the general questions in the first phase, as well as the timing, structure and questions of the second phase, are kept constant between the two treatments.

On the other hand, treatment T3 consists of a single phase, in which subjects first answer the general questions on their demographic, socio-economic and health conditions, and are then administered the same questions about their satisfaction with life domains and with life in general used in the second phase of the other two treatments.

Two aspects of our experimental design are particularly worth noting from a methodological perspective. First, the effect of enhancing awareness on life satisfaction is ascertained by disentangling the mere effect of providing information about the domains from that of letting subjects think deeply and evaluate each aspect of life. Second, our design allows us to assess results both between and within subjects, by comparing the distribution of life satisfaction in the first phase across the three treatments and between the two phases of treatments T1 and T2 respectively.

### 3.2. Procedures

The survey experiment took place between January and February 2013 and was administered by using Qualtrics (<http://www.qualtrics.com/>). The subjects were mainly students of economics from three different universities in northern Italy (namely Bocconi University in Milan, the University of Varese-Insubria and the University of Padova) and were recruited by e-mail after advertising the experiment through Facebook university groups. Once they had agreed to participate in the study, each subject was randomly, independently and anonymously assigned to (only) one of the three different treatments. To guarantee anonymity and to match the responses correctly across the two phases of treatments T1 and T2, subjects were required to provide the first six digits of their personal tax code (which is an alphanumeric code of 16 characters). To increase the external validity of our survey study, we followed the standard practice in well-known general surveys of not paying for subjects' participation.

### 3.3. Testable predictions

Two standard assumptions behind survey studies on the determinants of subjective wellbeing (see Schwarz and Strack (1999)) are that

- (a) subjects can evaluate their satisfaction with life and
- (b) their evaluations do not depend on the order in which the questions are asked, e.g. on whether the question on overall life satisfaction is asked before or after specific life domains are presented in the questionnaire.

This framework provides null predictions for our study, as we should observe differences neither in the distribution of life satisfaction evaluations across treatments nor in the correlation between evaluations of life domains and the measures of life satisfaction that are reported in the two phases of treatments T1 and T2.

However, as discussed in Section 2, there is a rich literature highlighting the existence of context effects in survey studies (see Schwarz and Strack (1999)). In this respect, we are interested in assessing how the questionnaire manipulations that were introduced in our survey experiment influence the observed distribution of life satisfaction and its association with life domains.

There are valid arguments to expect both the mean and the variance of the distribution of evaluations to vary across treatments and between phases. First, we expect the distribution of life satisfaction to change depending on whether and how much individuals are satisfied with each domain, and on how much weight is attached to each domain. Scholars of support theory suggest that, in evaluative judgements, 'the whole is less than the sum of its parts' (Van Boven and Epley (2003)) and priming important details of objects to evaluate might induce subjects to report more extreme evaluations on the ordered scale that is used to express responses. More specifically, both in expressing positive (for instance, anticipated enjoyment from a vacation in the Bahamas) and negative (for instance, suffering from health detriments from pollution produced by an oil refinery) evaluations, subjects tend to report higher values on the positive or negative scale under the unpacking manipulation: more informed people are less likely to express mild judgements. Since most of our sample is composed of relatively young, middle-class, well-educated and healthy individuals, we expect unpacking effects to affect life satisfaction evaluations positively in our sample. Furthermore, we expect that the mere effect of providing information on important aspects of life will be weaker than the more salient priming mechanism relying on evaluating each life domain.

Second, by affecting the accessibility and salience of information, priming life domains might facilitate subjects in building up a more complete image of life satisfaction (Schwarz and Strack, 1991, 1999; Schwarz, 1999). Thus, it is reasonable to expect the variance of responses to be influenced by the questionnaire manipulations, as subjects are likely to express more precise evaluations when information on life domains is provided, enhancing the reliability of life satisfaction. Again, as we manipulate the salience of the life domains across treatments and phases, we should expect the effect on the variance to be stronger when subjects express specific and separate evaluations on the domains than in the case in which they are simply presented with the list containing indications on these aspects of life.

Finally, we also want to assess how enhancing salience of the life domains affects the correlation between life satisfaction and the evaluation that is expressed for each domain, i.e. its validity as a proxy for overall wellbeing. As highlighted in the previous section, subjects in treatment T1 report their life satisfaction twice, with the second response (in the second phase) being expressed after the evaluations of the life domains. Thus, in formulating their second response, subjects in treatment T1 can use all the accessible information on life satisfaction that they obtain by facing specific questions on satisfaction with life domains (Schwarz and Strack, 1991, 1999). In addition to facilitating subjects to formulate a mental representation of life satisfaction, it is reasonable to expect the second response to exhibit a stronger (and more robust) association with the evaluations of the life domains just stated, thus making the unpacked questionnaire a more valid instrument to measure people's overall welfare. Of course, under the assumption that the salience of life domains depends on priming manipulations, and that listing the domains has a weaker priming effect than asking subjects to provide evaluations of each domain, we should observe an effect of priming on correlations between life satisfaction and evaluations of life domains in treatment T2.

To formalize these intuitions better, we develop a simple quantitative model. The model is very stylized: we treat the psychological processes underlying the unpacking effect as a black box, and the model aims only at representing the *outcomes* of these processes for the evaluation of life satisfaction. In particular, as stated above, according to the theory of unpacking (see Van Boven and Epley (2003)), the unawareness about the concept to be evaluated related to the packed questionnaire should lead people to concentrate their evaluations towards central (i.e. 'mild') values of the life satisfaction scale. We model this shrinkage by using a simple 'mean reverting' measurement error model, that we borrow from Bound and Krueger (1991). To keep



the model as simple as possible, we compare just a packed and an unpacked questionnaire, but the model can easily be extended to include an intermediate manipulation.

Suppose that true overall life satisfaction of individual  $i$ ,  $LS_i$ , is given by the weighted average of satisfaction in each of  $k = 1, \dots, K$  domains,  $LS_i^k$ ,

$$LS_i = \sum_{k=1}^K \beta^k LS_i^k. \quad (1)$$

For simplicity, we assume that the weights  $\beta^k$  are homogeneous across the population.  $LS_i$  is, however, unobserved, and all we can observe in the data is a reported measure of overall life satisfaction. We respectively refer to reported overall life satisfaction when the packed and unpacked questionnaire version are used as  $LS_i^p$  and  $LS_i^u$ . We suppose that individuals commit a systematic error in evaluating overall life satisfaction when they are asked about it by using the packed questionnaire version, so that  $LS_i^p$  takes the form

$$LS_i^p = LS_i + u_i, \quad (2)$$

$$\text{cov}(LS_i, u_i) < 0, \quad (3)$$

$$u_i \sim \text{IID}(0, \sigma^2) \quad (4)$$

where  $u_i$  is the measurement error, which is independently and identically distributed in the full population, with zero mean and variance  $\sigma^2$ . The negative covariance between  $u_i$  and true life satisfaction implies a mean reverting structure of reported satisfaction, which is shrunk towards milder central evaluations—as predicted by the unpacking theory. Given the well-documented positive relationship between satisfaction with life domains and overall life satisfaction, we also assume a negative correlation between the measurement error and domains' satisfaction:

$$\text{cov}(LS_i^k, u_i) < 0, \quad \forall k \quad (5)$$

In contrast, we assume that—by increasing awareness—the unpacked questionnaire manipulation reduces the amount of measurement error in the perception of satisfaction, so that  $LS_i^u$  can be expressed as

$$LS_i^u = LS_i + \lambda u_i, \quad \lambda < 1. \quad (6)$$

To keep the algebra as simple as possible, we assume that the weight on measurement error is 0 when the unpacked questionnaire is used, so that  $LS_i^u = LS_i$ . In this way, the model could be easily modified to add the intermediate manipulation, that provides only information, by setting  $0 < \lambda < 1$  for that manipulation.

We here present the testable predictions that we obtain from the model, whereas the proofs of the statements are reported in our on-line appendix. Our first testable prediction concerns the mean of reported life satisfaction in the unpacked and the packed version of the questionnaire in our positively selected sample of relatively young, middle-class, well-educated and healthy individuals. Because of the positive selection of our sample with respect to the full population, we have that

$$E(LS_i^p | \text{sample}) < E(LS_i^u | \text{sample}). \quad (7)$$

Our second testable prediction regards the variance of reported overall life satisfaction in the unpacked and the packed version of the questionnaire in our sample. We have that

$$\text{var}(LS_i^p | \text{sample}) = \text{var}(LS_i^u | \text{sample}) + \text{var}(u_i | \text{sample}) + 2 \text{cov}(LS_i, u_i | \text{sample}). \quad (8)$$

With respect to our initial hypothesis, the model adds an interesting insight about the effect of unpacking on the variance of reported life satisfaction. In fact, according to the model, unpacking decreases the overall variance of reported life satisfaction in our sample—making it a more reliable measure of wellbeing—only if the variance of the measurement error dominates its negative covariance with true life satisfaction:

$$\text{var}(u_i|\text{sample}) > -2\text{cov}(\text{LS}_i, u_i|\text{sample}).$$

Finally, our third testable prediction concerns the covariance between reported overall life satisfaction and satisfaction with life domains in our sample. Our model implies that

$$\text{cov}(\text{LS}_i^p, \text{LS}_i^k|\text{sample}) < \text{cov}(\text{LS}_i^u, \text{LS}_i^k|\text{sample}). \quad (9)$$

Hence, the covariance between reported overall life satisfaction and satisfaction with domains should be higher under the unpacked questionnaire manipulation, making this a more valid instrument to measure wellbeing.

#### 4. Empirical analysis

This section describes our empirical analysis. We start by presenting the data and by showing some descriptive statistics that are related to the testable predictions. Then we move to a more formal statistical analysis. We shall describe the empirical models exploited as we go through the analysis.

##### 4.1. Data and descriptive statistics

Our sample is composed of the subjects who took part in the experiment that was described in Section 2. To make sure that no changes in the underlying determinants of subjective wellbeing occurred between the two interviews, we drop the second-phase interviews of four subjects in treatments T1 and T2 who reported having experienced extraordinary life changes across the two phases, including negative—the loss of a close relative, job displacement and divorce—as well as positive—the birth of a child, a promotion, and so on—events. Information recalled through the questionnaire consists of a set of questions about satisfaction with life in general and with the six specific life domains presented in Section 2, plus standard sociodemographic controls. Aside from gender, from the raw data we generate a set of dummy variables for being younger than 30 years old, having a partner, having children, reporting very good or good health, having a college degree, family income below €16 000, between €16 000 and €30 000, between €30 000 and €56 000, above €56 000, meeting friends at least once a week, not taking part in any cultural, political, sport-related or religious association, and being born in northern Italy. Descriptive statistics for the full sample are reported in Table 1.

As a result of our sampling strategy, we end up selecting mainly students or young workers. Table 1 shows that about 80% of the sample is younger than 30 years old and that only 10% of the sample reports having a child. Close to 71% of subjects report being in good health, and 53% have completed a college degree, and figures on income categories show that most subjects come from middle-class or well-off families. Although the extrapolation of our results to the general population is not guaranteed, internal validity of our causal statements is granted by randomization of subjects across the three treatment groups. To test for balancing in sample composition across treatments, we regress each of the covariates that are presented in the upper panel of Table 1 on a constant and dummies for belonging to treatment T2 and T3 respectively. We report the constant and the coefficients for the two treatment group dummies in Table 2,

**Table 1.** Descriptive statistics†

<i>Covariate</i>	<i>Mean</i>	<i>Standard deviation</i>	<i>Observations</i>
Female	0.570	0.496	342
Age below 30 years	0.798	0.402	342
Has a partner	0.342	0.475	342
Has children	0.099	0.300	342
In good health	0.708	0.456	342
College degree	0.526	0.500	342
Mid-low income	0.307	0.462	342
Mid-high income	0.266	0.443	342
High income	0.170	0.376	342
Meeting friends often	0.678	0.468	342
Associations membership	0.459	0.499	342
From northern Italy	0.459	0.499	342
Satisfaction with life	7.371	1.470	469
Satisfaction with income	6.236	2.108	267
Satisfaction with family	7.562	1.894	267
Satisfaction with work or study	6.577	1.957	267
Satisfaction with friends	7.779	1.606	267
Satisfaction with partner	6.431	2.924	267
Satisfaction with health	7.790	1.639	267

†See Table 3 for sample size definition.

together with their Huber–White heteroscedasticity consistent standard errors and level of significance.

Overall, randomization worked well, as most differences are not statistically different from 0. However, we still detect some imbalances across the three groups, probably due to small sample size: for instance, subjects in treatment T3 are older and more likely to have children than subjects in treatment T1, whereas group T2 presents a higher percentage of female respondents than does group T1. Similar results are obtained when we estimate a generalized propensity score through a multinomial logit regression for treatment status on the same set of covariates (see Imbens (2000) and Lechner (2001)). The *pseudo-R*<sup>2</sup> of such a regression is around 0.1, confirming that the distribution of covariates across groups is only mildly different. To make sure that we remove any bias due to imbalances in the observable covariates, we are going to include them in our regressions. We also show that our regression results are unchanged whether we include or exclude covariates, confirming the reasonable balance arising from the randomization and supporting the internal validity of our findings.

Table 3 presents features of the distribution of reported overall life satisfaction, our dependent variable, and sample size by treatment and phase.

Evidence from Table 3 already provides useful insights about the effects of the treatment on the distribution on life satisfaction. We begin by focusing on results for phase 1. A comparison of outcomes of groups T1 and T2 suggests that providing a list of life domains that are relevant for satisfaction with life as a whole does not affect the distribution of overall satisfaction with life, either in terms of mean levels, or in terms of standard deviation within each group. In contrast, a comparison of the outcomes of groups T1 and T2 with group T3 reveals the first evidence of unpacking effects as well as increased awareness on the topic to be evaluated: asking subjects to rate satisfaction with specific life domains leads them to express more positive judgements and shrinks evaluations towards this higher mean level. Graphical evidence in this sense is reported in Fig. 2, where we plot the histogram of satisfaction with life across treatment groups in the

**Table 2.** Balancing tests<sup>†</sup>

<i>Covariate</i>	<i>(1) Mean, T1</i>	<i>(2) Mean, T2–T1</i>	<i>(3) Mean, T3–T1</i>
Female	0.495	0.149 <sup>‡</sup> (0.069)	0.076 (0.065)
Age below 30 years	0.911	–0.020 (0.042)	–0.261 <sup>§</sup> (0.049)
Has a partner	0.386	–0.010 (0.069)	–0.100 (0.062)
Has children	0.050	–0.020 (0.028)	0.136 <sup>§</sup> (0.039)
In good health	0.723	–0.059 (0.065)	0.006 (0.059)
College degree	0.396	0.050 (0.070)	0.283 <sup>§</sup> (0.063)
Mid–low income	0.277	0.069 (0.065)	0.023 (0.059)
Mid–high income	0.347	–0.129 <sup>‡</sup> (0.063)	–0.104 <sup>§§</sup> (0.060)
High income	0.178	–0.079 (0.049)	0.036 (0.052)
Meeting friends often	0.782	–0.079 (0.062)	–0.196 <sup>§</sup> (0.059)
Associations membership	0.455	–0.010 (0.070)	0.016 (0.065)
From northern Italy	0.584	–0.069 (0.070)	–0.256 <sup>§</sup> (0.063)
Observations	342		

<sup>†</sup>We report mean values of the covariates for individuals in treatment group 1 in column (1) and differences in mean values between treatment groups 2 and 1, and between treatment group 3 and 1 in columns (2) and (3) respectively. Huber–White heteroscedasticity consistent standard errors are given in parentheses.

<sup>‡</sup> $p < 0.05$ .

<sup>§</sup> $p < 0.01$ .

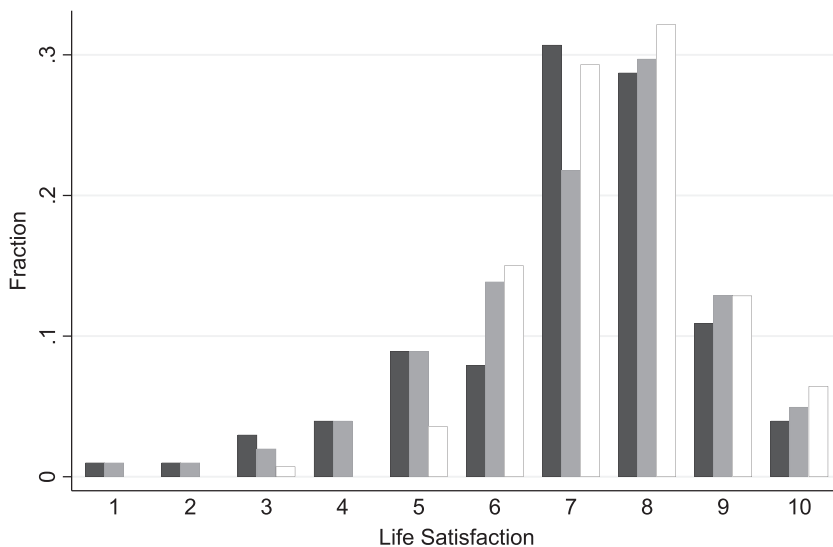
<sup>§§</sup> $p < 0.1$ .

full sample of phase 1. To test for differences in the distribution of life satisfaction between the treatments, we use the non-parametric Epps–Singleton two-sample test. The test detects no significant differences between groups T1 and T2 ( $p$ -value 0.978), whereas the distribution of life satisfaction of group T3 is significantly different from that obtained by pooling groups T1 and T2 ( $p$ -value 0.0028). Fig. 2 also shows that our result on the variance is unlikely owing to a ceiling effect (i.e. truncation of the scale from above), as only a small fraction of subjects report evaluations of satisfaction with life by using the highest available point on the evaluation scale.

Finally, similar evidence of the unpacking effect is also present when we consider the longitudinal dimension of our experiment and compare the distribution of satisfaction with life that subjects in groups T1 and T2 express in the two phases. In this sense, it is worth noting that around 40% of the initial T1 and T2 subjects drop out from the survey between the two phases. As a consequence, longitudinal findings might be biased because of panel attrition if only people reporting higher satisfaction with life remain in the sample. We test for endogenous attrition by comparing mean baseline characteristics of the full sample and of the sample of ‘stayers’, both separately for groups T1 and T2 and jointly. The idea is that we want to test whether the sample of stayers is representative of the initial population, as suggested by Fitzgerald *et al.*

**Table 3.** Life satisfaction across treatments and phases

Phase	Results for treatment 1	Results for treatment 2	Results for treatment 3
1	Mean: 7.020 Standard deviation: 1.726 Observations: 101	Mean: 7.079 Standard deviation: 1.747 Observations: 101	Mean: 7.521 Standard deviation: 1.249 Observations: 140
2	Mean: 7.655 Standard deviation: 1.001 Observations: 58	Mean: 7.768 Standard deviation: 1.178 Observations: 69	

**Fig. 2.** Histogram of life satisfaction in phase 1, by treatment group: ■, treatment 1 (packed); ■, treatment 2 (information); □, treatment 3 (unpacked)

(1998). Results are presented in Table 4 and do not reveal any evidence of endogenous attrition, as we find no significant imbalance between the two groups either in terms of the observable covariates or of the evaluation of life satisfaction carried out at baseline.

#### 4.2. Statistical analysis

We carry out formal statistical analyses in this section. We start by exploiting the variation in treatment assignment between groups within the first phase of the experiment and analyse the effects of the questionnaire manipulations on the mean and the variance of overall satisfaction with life. To estimate treatment effects on the mean of the dependent variable, we run simple linear regressions of overall life satisfaction on dummies for treatments T2 and T3 and the set of covariates that were illustrated in the previous section, using Huber–White heteroscedasticity consistent standard errors. However, to estimate treatment effects on the unconditional variance of satisfaction with life across the three treatment groups, we exploit recentred influence function regressions, as introduced by Firpo *et al.* (2009). The idea of the method is to replace the dependent variable by the corresponding recentred influence function for the distributional statistics of interest (Clark *et al.* (2014) also exploited recentred influence function regressions

**Table 4.** Attrition†

<i>Covariate</i>	<i>Sample</i>	<i>Mean, full sample</i>	<i>Mean, stayers</i>	<i>Difference</i>
Satisfaction with life	T1 and T2	7.050	7.205	−0.155
	T1	7.020	7.069	−0.049
	T2	7.079	7.319	−0.240
Female	T1 and T2	0.569	0.559	0.010
	T1	0.495	0.500	−0.005
	T2	0.644	0.609	0.035
Age below 30 years	T1 and T2	0.901	0.921	−0.020
	T1	0.911	0.966	−0.055
	T2	0.891	0.884	0.007
Has a partner	T1 and T2	0.381	0.378	0.003
	T1	0.386	0.397	−0.010
	T2	0.376	0.362	0.014
Has children	T1 and T2	0.040	0.024	0.016
	T1	0.050	0.017	0.032
	T2	0.030	0.029	0.001
In good health	T1 and T2	0.693	0.717	−0.023
	T1	0.723	0.724	−0.001
	T2	0.663	0.710	−0.047
College degree	T1 and T2	0.421	0.433	−0.012
	T1	0.396	0.397	−0.001
	T2	0.446	0.464	−0.018
Mid–low income	T1 and T2	0.312	0.307	0.005
	T1	0.277	0.207	0.070
	T2	0.347	0.391	−0.045
Mid–high income	T1 and T2	0.282	0.315	−0.033
	T1	0.347	0.448	−0.102
	T2	0.218	0.203	0.015
High income	T1 and T2	0.139	0.126	0.013
	T1	0.178	0.155	0.023
	T2	0.099	0.101	−0.002
Meeting friends often	T1 and T2	0.743	0.787	−0.045
	T1	0.782	0.828	−0.045
	T2	0.703	0.754	−0.051
Associations membership	T1 and T2	0.450	0.433	0.017
	T1	0.455	0.448	0.007
	T2	0.446	0.420	0.025
From northern Italy	T1 and T2	0.550	0.543	0.006
	T1	0.584	0.552	0.032
	T2	0.515	0.536	−0.021

†The sample considered includes treatment groups 1 and 2. We report mean values of life satisfaction and the covariates at baseline in the full sample and in the sample of those who do not drop out between phase 1 and phase 2, and the difference in the means.

on the variance of satisfaction with life). In both cases, identification is granted by random assignment to treatment, as documented in Section 4.1. Inclusion of covariates enforces conditional independence in case of randomization failure and helps to increase precision of our estimates. Results are presented in Table 5.

Looking at column (1), we see that presenting respondents with information on specific domains that are relevant for overall satisfaction with life before expressing a general judgement does not significantly affect evaluations relative to the benchmark treatment T1. In contrast, asking respondents to elicit satisfaction with specific life domains before the general question exerts a significant and strongly positive unpacking effect on satisfaction with life: we detect a mean shift of 0.502–0.627, depending on the empirical specification. This shift corresponds

**Table 5.** Unpacking life satisfaction†

<i>Covariate</i>	<i>(1) Mean</i>	<i>(2) Mean</i>	<i>(3) Variance</i>	<i>(4) Variance</i>
T2—information	0.282 (0.221)	0.0594 (0.244)	−0.313 (0.657)	0.0504 (0.654)
T3—unpacked	0.627‡ (0.203)	0.502§ (0.202)	−1.923‡ (0.670)	−1.373§ (0.607)
Female	−0.00609 (0.157)		0.569 (0.522)	
Age below 30 years	0.823‡ (0.245)		−1.737§ (0.774)	
Has a partner	0.165 (0.170)		0.458 (0.578)	
Has children	0.851§ (0.343)		−1.189 (1.071)	
In good health	0.960‡ (0.189)		−1.646‡ (0.552)	
College degree	0.486‡ (0.173)		−0.596 (0.555)	
Mid—low income	0.372§§ (0.209)		−1.588§ (0.663)	
Mid—high income	0.583§ (0.235)		−0.933 (0.696)	
High income	0.687‡ (0.254)		−0.148 (0.795)	
Meeting friends often	0.630‡ (0.185)		−1.164§ (0.570)	
Associations membership	0.00714 (0.152)		−0.0543 (0.506)	
From northern Italy	0.00739 (0.172)		−0.611 (0.541)	
Constant	4.353‡ (0.469)	7.020‡ (0.172)	7.689‡ (1.246)	3.000‡ (0.462)
Observations	342	342	342	342
$R^2$	0.234	0.022	0.104	0.022

†The dependent variable is overall satisfaction with life. Columns (1) and (2) report the ordinary least squares regression coefficients. Columns (3) and (4) report the recentred influence function regression coefficients for the variance of life satisfaction. The results refer to phase 1 only. Huber–White heteroscedasticity robust standard errors are given in parentheses.

‡ $p < 0.01$ .

§ $p < 0.05$ .

§§ $p < 0.1$ .

to 7.2–8.9% of the mean of life satisfaction in the untreated (T1) group, that is equal to 7.02. These differences are consistent with those reported in Table 3 and with the predictions of our quantitative model. Indeed, taking the model at face value, the treatment effect between treatments T1 and T3 on the mean of reported overall life satisfaction identifies the (opposite of the) mean of the error term  $u_i$  in the selected sample of respondents.

*Result 1.* In our sample, evaluating life domains before overall life satisfaction increases the mean of overall life satisfaction.

To assess the robustness of our results to the parametric specification that is imposed by the model, we also consider an alternative semiparametric estimator. We implement the multivalued treatment propensity score weighting estimator that was discussed in Imbens (2000) and Lechner

(2001). We first estimate the probability of receiving the treatment actually received—the so-called ‘propensity score’—through a (parametric) multinomial logit model, and we then (non-parametrically) compare the average outcomes by treatment group, weighted by the inverse of the propensity score. In this way, we both test that no covariate imbalancing is present after weighting each observation for the propensity score, and that, when we consider reported overall life satisfaction as our dependent variable, estimation results from this different specification—not shown—are quantitatively and qualitatively equivalent to our baseline model. Results are also robust to dropping observations that are extreme with respect to our propensity score metric. Finally, equivalent results—not shown but available on request—are obtained when we treat overall life satisfaction as an ordinal measure and use ordered probit models.

We now turn our attention to understanding how adding explicit references to the life domains affects the variance of the reported levels of life satisfaction. As shown in column (3) of Table 5, we find that simply presenting subjects with a list containing the life domains does not influence the standard deviation of life satisfaction, whereas asking subjects to elicit their satisfaction with each specific domain exerts a strong negative effect on the variance of life satisfaction (columns (2) and (4) show that results are not dependent on the inclusion of covariates). We estimate that the unpacked treatment decreases the variance of life satisfaction by from  $-1.923$  to  $-1.373$ , depending on the specification, which corresponds to a decrease by from  $-64\%$  to  $-46\%$  of the variance of life satisfaction in the untreated (T1) group. These differences are also consistent with those presented in Table 3 in terms of standard deviations. With respect to the predictions of our quantitative model, these results highlight that the variance of the measurement error in the packed questionnaire dominates its negative covariance with true life satisfaction, since the variance of reported overall life satisfaction is lower by using the unpacked rather than the packed questionnaire version (see equation (8)).

*Result 2. Evaluating life domains reduces the variance of the distribution of life satisfaction.*

By raising awareness of the domains to be evaluated, asking subjects to report their satisfaction with the six life domains leads them to take a less mild position on their satisfaction with life and to reduce uncertainty in the evaluations, which shrink towards a higher mean level. Furthermore, focusing on the estimates that are presented in column (1), most coefficients that are related to the covariates included in the equation have the expected signs (see Frey and Stutzer (2002a, b) and Dolan *et al.* (2008)): satisfaction increases with income, and it is higher for the youngest in the sample, for those with children, those in good health and those who have more frequent contacts with friends. As a robustness check, we have also repeated this analysis on the subsample of stayers only, and results—not shown for brevity—are unchanged.

Next, we exploit the variation between phases and within treatment by taking advantage of the fact that, in both group T1 and group T2, subjects state their life satisfaction twice, with the second evaluation being expressed after judging satisfaction with the six life domains. This longitudinal set-up allows us to estimate the effects of life domain evaluations on life satisfaction by using within-subject variation. Since no difference in life satisfaction at baseline was detected between treatments T1 and T2, we pool observations from these two groups and include a dummy variable for belonging to group T2 in all models (we also checked for heterogeneous effects, but the interaction term was not significant). Since we consider repeated observations for the same individual, standard errors take into account the clustering within individuals. Table 6 presents estimation outcomes.

Results on the mean and the variance of overall life satisfaction are consistent with our previous findings both qualitatively and quantitatively: when subjects are primed with their own judgements on specific life domains, the mean level of life satisfaction increases and the



**Table 6.** Longitudinal analysis

	(1) Mean	(2) Mean	(3) Variance	(4) Variance
Second phase	0.512 <sup>‡</sup> (0.146)	0.512 <sup>‡</sup> (0.143)	-1.786 <sup>‡</sup> (0.573)	-1.786 <sup>‡</sup> (0.591)
Covariates	Yes	No	Yes	No
Observations	254	254	254	254
R <sup>2</sup>	0.210	0.034	0.135	0.035

<sup>†</sup>The dependent variable is overall satisfaction with life. Columns (1) and (2) report the mean ordinary least squares regression coefficient associated with the second-phase dummy. Columns (3) and (4) report the recentred influence function regression coefficient for the variance of life satisfaction. The sample considered includes only individuals observed in both phases. The covariates used in columns (1) and (3) are shown in the upper panel of Table 1. A dummy for treatment group 2 is also included. Standard errors that take into account the clustering within individuals are reported in parentheses.

<sup>‡</sup> $p < 0.01$ .

**Table 7.** Not a retesting effect?<sup>†</sup>

	(1) Mean	(2) Mean	(3) Variance	(4) Variance
T2—information	0.215 (0.201)	0.113 (0.193)	0.265 (0.369)	0.404 (0.365)
T3—unpacked	-0.036 (0.197)	-0.134 (0.168)	0.187 (0.363)	0.572 <sup>‡</sup> (0.320)
Covariates	Yes	No	Yes	No
Observations	267	267	267	267
R <sup>2</sup>	0.170	0.008	0.109	0.012

<sup>†</sup>The dependent variable is overall satisfaction with life. Columns (1) and (2) report the mean ordinary least squares regression coefficients associated with the treatment dummies. Columns (3) and (4) report the recentred influence function regression coefficients for the variance of life satisfaction. Covariates included in columns (1) and (3) are shown in the upper panel of Table 1. The sample considered includes only the treatment groups in which domains are elicited. Huber–White heteroscedasticity consistent standard errors are given in parentheses.

<sup>‡</sup> $p < 0.1$ .

distribution of the measure becomes more concentrated around this higher value. Furthermore, results are robust to the exclusion of individual covariates, confirming the robustness of our findings.

As a placebo test, we also compared the distribution of the reported levels of satisfaction with life expressed in phase 2 by subjects in treatments T1 and T2 with those expressed by subjects in treatment T3 in phase 1. Finding no differences in the mean and the variance of the distribution of life satisfaction across treatment groups does not allow us to conclude that our results in the longitudinal analysis are not due to retesting effects, because we never observe a group receiving no information in both phases, yet it is reassuring to see that results in Table 7 confirm that no difference across groups that are exposed to the same treatment in different phases is detectable (furthermore, to the best of our knowledge, no evidence of positive changes in life satisfaction due to retesting is present in the literature).

**Table 8.** Overall life satisfaction and satisfaction with domains†

	(1) Results for phase 1	(2) Results for phase 1	(3) Results for phase 2	(4) Results for phase 2	(5) Results for T3	(6) Results for T3
Satisfaction with income	0.026 (0.081)	0.050 (0.104)	0.103‡ (0.039)	0.104‡ (0.038)	0.118‡ (0.036)	0.117‡ (0.035)
Satisfaction with family	0.131 (0.081)	0.140§ (0.084)	0.159‡ (0.039)	0.167‡ (0.037)	0.214‡ (0.033)	0.215‡ (0.033)
Satisfaction with work or study	0.069 (0.099)	0.144 (0.103)	0.107§§ (0.050)	0.110§§ (0.047)	0.163‡ (0.033)	0.163‡ (0.033)
Satisfaction with friends	0.137 (0.103)	0.136 (0.111)	0.149‡ (0.053)	0.152‡ (0.051)	0.101‡ (0.035)	0.101‡ (0.035)
Satisfaction with partner	0.124§§ (0.056)	0.075 (0.060)	0.048§ (0.027)	0.046§ (0.026)	0.077‡ (0.025)	0.077‡ (0.025)
Satisfaction with health	0.093 (0.085)	0.074 (0.088)	0.165‡ (0.054)	0.162‡ (0.053)	0.192‡ (0.037)	0.191‡ (0.036)
Covariates	Yes	No	Yes	No	Yes	No
Observations	127	127	127	127	140	140
R <sup>2</sup>	0.261	0.174	0.594	0.591	0.675	0.674

†The dependent variable is overall satisfaction with life, and we report mean ordinary least squares regression coefficients associated with satisfaction with specific domains. The sample considered includes only individuals who are observed for two phases. Columns (1) and (2) consider outcomes for phase 1, columns (3) and (4) for phase 2 and columns (5) and (6) for the third treatment. Covariates included in columns (1) and (3) are age, geographical origin and gender. Huber–White heteroscedasticity consistent standard errors are given in parentheses.

‡ $p < 0.01$ .

§ $p < 0.1$ .

§§ $p < 0.05$ .

So far, our results suggest that asking subjects to evaluate specific life domains before expressing an overall judgement on satisfaction with life raises their awareness about the general topic to be assessed and allows them to express more accurate evaluations. As a consequence, we expect to observe a higher correlation between life satisfaction and satisfaction with life domains when these are elicited before the general question. To test this hypothesis, we focus on subjects in the longitudinal sample, for whom we observe two general evaluations: one expressed without prior assessment of specific life domains, in phase 1, and one elicited after domains evaluation, in phase 2. We run two simple linear regressions of the two overall evaluations on the evaluations of the specific domains and controls for gender, age and geographical origin. We drop other covariates as we consider that they measure concepts which are close to those captured by the specific domains. Results are presented in Table 8 and suggest that evaluations carried out in phase 2, after domains elicitation, are more strongly correlated with satisfaction with the domains. We also formally test whether the estimated coefficients for phase 1 and phase 2 are significantly different through a Chow test (Chow, 1960). The test confirms that these differences are statistically significant (the  $p$ -value is 0.010 when including the restricted set of covariates and 0.065 when excluding them). Furthermore, the  $R^2$ -value of regressions for phase 1 is equal to 0.26 and 0.17 with and without covariates respectively, whereas it equals 0.59 for both phase 2 regressions. Confirming the predictions of our quantitative model, these findings show that raising awareness about life domains decreases measurement error in the general evaluations, thus increasing the coherence between self-reported satisfaction with domains and satisfaction with life as a whole. As a robustness check, we repeat the same analysis for treatment T3 at phase 1, where domain evaluations are elicited before asking the general question. As expected, the results are comparable with those of phase 2 (see columns (5) and (6) of Table 8).

**Table 9.** Interaction effects†

	<i>Life satisfaction</i>
Second phase	2.310‡ (0.677)
Second phase * Female	0.289 (0.249)
Second phase * Age below 30 years	-1.230‡ (0.355)
Second phase * Has a partner	0.110 (0.306)
Second phase * Has children	-1.412‡ (0.283)
Second phase * In good health	-0.708§ (0.344)
Second phase * College degree	0.0923 (0.277)
Second phase * Mid-low income	0.0378 (0.382)
Second phase * Mid-high income	0.142 (0.392)
Second phase * High income	-0.140 (0.518)
Second phase * Meeting friends often	-0.540 (0.383)
Second phase * Associations membership	-0.524§ (0.263)
Second phase * From northern Italy	0.452 (0.315)
Observations	254
$R^2$	0.270

†The dependent variable is overall life satisfaction. All covariates for which interaction effects are shown are included linearly as well. A dummy for belonging to treatment group 2 is also included. We consider the longitudinal sample only and we estimate the model by ordinary least squares. Standard errors that take into account the clustering within individuals are reported in parentheses.

‡ $p < 0.01$ .

§ $p < 0.05$ .

*Result 3.* Evaluating life domains increases the correlation between these responses and the (subsequently) reported level of life satisfaction.

We now turn to a subgroup analysis, aimed at assessing whether the context effects that we have detected are homogeneous across the population. Assuming that context effects affect everyone in the same way, as first stressed by Bertrand and Mullainathan (2001), the main risk related to using measures of life satisfaction from questions asked ‘out of the blue’ within a survey is related to the loss of estimation precision if these measures are used as outcomes, or attenuation bias if these measures are used as controls. The risks could even be more worrisome, however. If the effects of using contextualized *vis-à-vis* uncontextualized questions are heterogeneous with respect to individual characteristics such as age, gender or health status, then coefficients related to these variables in a subjective wellbeing regression will change depending on the questionnaire version that is used, making it difficult for analysts to understand whether subgroup differences in wellbeing are germane or due to context effects.

Few references have analysed the heterogeneous effect of context effects on life satisfaction evaluations across different subgroups of respondents. For instance, Deaton and Stone (2013) compared how the self-reported levels of wellbeing change when individuals are formerly asked to respond to a series of political questions. They found that elderly respondents report higher life evaluation than young respondents when the political questions are asked, but lower life evaluation in their absence. Moreover, African-Americans report higher life evaluations than white respondents when they are asked about President Obama's performance. By using a more similar setting to ours, McClendon and O'Brien (1988) found that married respondents report higher life satisfaction when formerly asked a set of questions on satisfaction in eight specific life domains. In contrast, no order effects were found for unmarried people.

Even if our sample is quite homogeneous in terms of observable characteristics (see Table 1), the analysis on our longitudinal sample, which is presented in Table 9, shows that there is indeed heterogeneity in the effect of the unpacking manipulation on the basis of individuals' baseline characteristics. When we add interaction terms between the second-phase dummy and individual level covariates, we find that the phase 2 dummy is still significant and that the interactions are also jointly significant ( $p$ -value less than 0.001). Hence, we reject that the unpacking effect is homogeneous across all people in our sample. The coefficients that are related to the interaction terms also convey interesting evidence. For instance, the youngest, parents and people in good health at baseline experience smaller changes in life satisfaction when exposed to the unpacked questionnaire than the oldest, those without children and people in bad health. The unpacking manipulation may allow the latter groups to consider other life domains beyond, for instance, their bad health when assessing overall satisfaction with life, with a positive effect on their self-reported wellbeing level. This interpretation is in line with the well-known 'focusing illusion' effect (see Schkade and Kahneman (1998) and Wilson *et al.* (2000)), according to which people tend to overstate the emotional consequences of a particularly salient circumstance, neglecting the overall context in which this takes place. Hence, this effect is also attenuated by the use of an unpacked version of the life satisfaction question.

## 5. Conclusions

The aim of the paper was to assess how raising awareness of six specific life domains—income, family, job, friends, sentimental relationships and health—affects subjects' self-reported levels of life satisfaction. To investigate the relevance of this specific context effect, we administered a survey experiment based on three different questionnaire versions that can be ordered on the basis of the intensity of the awareness manipulation: one with no reference to the life domains, one including simply the list of the life domains and, finally, one in which subjects evaluate each domain before expressing their overall satisfaction with life.

We document a strong and robust unpacking effect, whereby evaluating the (subjective) conditions with the domains makes subjects in our sample more satisfied with their life. In addition, raising awareness of the domains substantially reduces the dispersion of responses and increases the association between life satisfaction and life domain evaluations. Finally, we have shown that the effects are not homogeneous across subgroups, with relevant implications for the estimation of subjective wellbeing equations by using different versions of questionnaire.

Our study is not exempted from limitations that are mainly related to the application to more realistic settings and the generalizability of the conclusions. Of course, these considerations suggest the necessity to replicate and extend our approach. Nevertheless, we believe that the contribution of the present paper to the existing literature is twofold. First, our results offer relevant insights into the flourishing empirical literature on life satisfaction. They suggest that

framing and context effects play a substantial role in isolating what people should refer to when asked to evaluate their satisfaction with life and stress the necessity to raise awareness of the determinants of life satisfaction to produce more precise and reliable subjective measures of wellbeing. Second, our results also caution researchers about the potential problems of comparability of data across different sources. Whereas in some cases the overall evaluation of life satisfaction is anticipated by questions concerning the subjective conditions with specific life domains (see for instance the 15th wave of the British Household Panel Survey, the 2004 edition of the German Socio-Economic Panel and the second wave of the Household, Income and Labor Dynamics in Australia survey), in other cases life satisfaction is assessed with no reference to the life domains (to mention some examples, the sixth wave of the European Social Survey and the sixth wave of the World Value Survey). As suggested by the present study, manipulating the structure of the questionnaire makes life satisfaction assessments not entirely comparable across different surveys and waves, as it is likely to alter the distribution of the responses substantially.

## Acknowledgements

The survey experiment was run by Alessandra Cometa for her Master of Science thesis in Management at Bocconi University under the supervision of Luca Corazzini. We thank Alessandra for her brilliant research assistance. We are grateful to the Editors, two referees, Abel Brodeur, Martina Celidoni, Conchita D'Ambrosio, Michele De Nadai, Jochen Mierau, Mirko Moro, Enrico Rettore, Lorenzo Rocco and to the audiences at seminars in Aarhus, Aix-en-Provence, Groningen and Verona for useful comments. The usual disclaimers apply.

## References

- Angelini, V., Cavapozzi, D., Corazzini, L. and Paccagnella, O. (2014) Do Danes and Italians rate life satisfaction in the same way?: Using vignettes to correct for individual-specific scale biases. *Oxf. Bull. Econ. Statist.*, **76**, 643–666.
- Bateman, I., Munro, A., Rhodes, B., Starmer, C. and Sugden, R. (1997) Does part-whole bias exist?: An experimental investigation. *Econ. J.*, **107**, 322–332.
- Benjamin, D. J., Kimball, M. S., Heffetz, O. and Rees-Jones, A. (2012) What do you think would make you happier?: What do you think you would choose? *Am. Econ. Rev.*, **102**, 2083–2110.
- Bernasconi, M., Corazzini, L., Kube, S. and Maréchal, M. A. (2009) Two are better than one!: Individuals' contributions to unpacked public goods. *Econ. Lett.*, **104**, 31–33.
- Bertrand, M. and Mullainathan, S. (2001) Do people mean what they say?: Implications for subjective survey data. *Am. Econ. Rev.*, **91**, 67–72.
- Bound, J. and Krueger, A. B. (1991) The extent of measurement error in longitudinal earnings data: do two wrongs make a right? *J. Lab. Econ.*, **9**, 1–24.
- Bruni, L. and Porta, P. L. (2005) *Economics and Happiness: Framing the Analysis*. Oxford: Oxford University Press.
- Chow, G. C. (1960) Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, **28**, 591–605.
- Clark, A. E., Etilé, F., Postel-Vinay, F., Senik, C. and Van der Straeten, K. (2005) Heterogeneity in reported well-being: evidence from twelve European countries. *Econ. J.*, **115**, C118–C132.
- Clark, A. E., Flèche, S. and Senik, C. (2014) The great happiness moderation. In *Happiness and Economic Growth: Lessons from Developing Countries* (eds A. E. Clark and C. Senik), pp. 32–139. Oxford: Oxford University Press.
- Clark, A. E., Frijters, P. and Shields, M. A. (2008) Relative income, happiness, and utility: an explanation for the Easterlin paradox and other puzzles. *J. Econ. Lit.*, **46**, 95–144.
- Conti, G. and Pudney, S. (2011) Survey design and the analysis of satisfaction. *Rev. Econ. Statist.*, **93**, 1087–1093.
- Cozby, P. C. (2001) *Measurement Concepts: Methods in Behavioral Research*. Mayfield.
- Cronbach, L. J. (1971) Test validation. In *Educational Measurement* (ed. R. L. Thorndike), pp. 443–507. Washington DC: American Council on Education.
- Deaton, A. (2012) The financial crisis and the well-being of Americans. *Oxf. Econ. Pap.*, **64**, 1–26.
- Deaton, A. and Stone, A. (2013) Do context effects limit the usefulness of self-reported wellbeing measures? *Working Paper 288*. Research Program in Development Studies, Princeton.

- Diener, E. (1984) Subjective well-being. *Psychol. Bull.*, **95**, 542–575.
- Diener, E., Napa-Scollon, C. K., Oishi, S., Dzokoto, V. and Suh, E. M. (2000) Positivity and the construction of life satisfaction judgments: global happiness is not the sum of its parts. *J. Happiness Stud.*, **1**, 159–176.
- Di Tella, R. and MacCulloch, R. (2006) Some uses of happiness data in economics. *J. Econ. Perspect.*, **20**, 25–46.
- Dolan, P., Peasgood, T. and White, M. (2008) Do we really know what makes us happy?: A review of the economic literature on the factors associated with subjective well-being. *J. Econ. Psychol.*, **29**, 94–122.
- Easterlin, R. A. (2001) Income and happiness: towards a unified theory. *Econ. J.*, **111**, 465–484.
- Ferrer-i Carbonell, A. and Frijters, P. (2004) How important is methodology for the estimates of the determinants of happiness? *Econ. J.*, **114**, 641–659.
- Firpo, S., Fortin, N. M. and Lemieux, T. (2009) Unconditional quantile regressions. *Econometrica*, **77**, 953–973.
- Fitzgerald, J., Gottschalk, P. and Moffitt, R. (1998) An analysis of sample attrition in panel data: The Michigan Panel Study of Income Dynamics. *J. Hum. Resour.*, **33**, 251–299.
- Frey, B. S. and Stutzer, A. (2002a) The economics of happiness. *Wrlld Econ.*, **3**, 25–41.
- Frey, B. S. and Stutzer, A. (2002b) What can economists learn from happiness research? *J. Econ. Lit.*, **40**, 402–435.
- Imbens, G. W. (2000) The role of the propensity score in estimating dose-response functions. *Biometrika*, **87**, 706–710.
- Kogut, T. and Ritov, I. (2005) The identified victim effect: an identified group, or just a single individual? *J. Behav. Decsn Makng*, **18**, 157–167.
- Lechner, M. (2001) Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. In *Econometric Evaluation of Labour Market Policies* (eds M. Lechner and F. Pfeiffer), pp. 43–58. Heidelberg: Physica.
- Lucas, R. E. and Lawless, N. M. (2013) Does life seem better on a sunny day?: Examining the association between daily weather conditions and life satisfaction judgments. *J. Personlty Socl Psychol.*, **104**, 872–884.
- McClendon, M. J. and O'Brien, D. J. (1988) Question-order effects on the determinants of subjective well-being. *Publ. Opin. Q.*, **52**, 351–364.
- van Praag, B. M. S. and Ferrer-i Carbonell, A. (2004) *Happiness Quantified: a Satisfaction Calculus Approach*. Oxford: Oxford University Press.
- Schkade, D. A. and Kahneman, D. (1998) Does living in California make people happy?: A focusing illusion in judgments of life satisfaction. *Psychol. Sci.*, **9**, 340–346.
- Schwarz, N. (1999) Self-reports: how the questions shape the answers. *Am. Psychol.*, **54**, 93–105.
- Schwarz, N. and Clore, G. L. (1983) Mood, misattribution, and judgments of well-being: informative and directive functions of affective states. *J. Personlty Socl Psychol.*, **45**, 513–523.
- Schwarz, N. and Strack, F. (1991) Evaluating one's life: a judgment model of subjective well-being. In *Subjective Well-being: an Interdisciplinary Perspective* (eds F. Strack, M. Argyle and N. Schwarz), pp. 27–47. Oxford: Pergamon.
- Schwarz, N. and Strack, F. (1999) Reports of subjective well-being: judgmental processes and their methodological implications. In *Well-being: the Foundations of Hedonic Psychology* (eds D. Kahneman, E. Diener and N. Schwarz), pp. 61–84. New York: Russell Sage Foundation.
- Schwarz, N., Strack, F. and Mai, H.-P. (1991) Assimilation and contrast effects in part-whole question sequences: a conversational logic analysis. *Publ. Opin. Q.*, **55**, 3–23.
- Small, D. A. and Loewenstein, G. (2003) Helping a victim or helping the victim: altruism and identifiability. *J. Risk Uncertainty*, **26**, 5–16.
- Small, D. A. and Loewenstein, G. (2005) The devil you know: the effects of identifiability on punishment. *J. Behav. Decsn Makng*, **18**, 311–318.
- Strack, F., Martin, L. L. and Schwarz, N. (1988) Priming and communication: social determinants of information use in judgments of life satisfaction. *Eur. J. Socl Psychol.*, **18**, 429–442.
- Strack, F., Schwarz, N. and Gschneidinger, E. (1985) Happiness and reminiscing: the role of time perspective, affect, and mode of thinking. *J. Personlty Socl Psychol.*, **49**, 1460–1469.
- Tourangeau, R. and Rasinski, K. A. (1988) Cognitive processes underlying context effects in attitude measurement. *Psychol. Bull.*, **103**, 299–314.
- Tourangeau, R., Rasinski, K. A. and Bradburn, N. (1991) Measuring happiness in surveys: a test of the subtraction hypothesis. *Publ. Opin. Q.*, **55**, 255–266.
- Tversky, A. and Koehler, D. J. (1994) Support theory: a nonextensional representation of subjective probability. *Psychol. Rev.*, **101**, 547–567.
- Van Boven, L. and Epley, N. (2003) The unpacking effect in evaluative judgments: when the whole is less than the sum of its parts. *J. Exptl Socl Psychol.*, **39**, 263–269.
- Wilson, T. D., Wheatley, T., Meyers, J. M., Gilbert, D. T. and Axson, D. (2000) Focalism: a source of durability bias in affective forecasting. *J. Personlty Socl Psychol.*, **78**, 821–836.

#### Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Unpacking the determinants of life satisfaction: a survey experiment'.