

Supporting Information

Weiss et al. 10.1073/pnas.1212592109

SI Text

Table S1 describes, for each sample, the interrater reliabilities of the mean scores across raters of the individual well-being items and the composite well-being score. Interrater reliabilities estimates were defined as intraclass correlation formula $ICC(3,k)$, which is defined as the proportion of the variance between subjects that is true score variance (1). The estimates were derived from all subjects in each sample that were rated by more than one individual. We used standard guidelines (2) to interpret the reliability of ratings. None of the items in our three samples had poor reliability [$ICC(3,k) < 0.4$] and, in all but one instance (the reliability of asking how successful an orangutan was in achieving its goals was fair), the reliabilities were good [$ICC(3,k) = 0.60$ – 0.74] or excellent [$ICC(3,k) >$

0.74]. The reliabilities of the well-being composites were high (Table S1).

Table S2 is included as a general robustness check. It examines the appropriateness of fitting the shape discussed in the human well-being literature, namely a quadratic, to the full ape dataset. To do this check, the analysis presented in Table S2 estimates a well-being equation without imposing any parameterized structure or polynomial function. The results reveal that, even with an elementary set of 11 banded dummy variables, the low point is reached between age 30 and age 35, and that, although subsample sizes are inevitably too small within each age band to allow precision on individual coefficients or a perfect nonparametric U, there is evidence broadly consistent with the study's parameterized approach.

1. Shrout PE, Fleiss JL (1979) Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull* 86(2):420–428.

2. Cicchetti DV (1994) Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* 6:284–290.

Table S1. Interrater reliabilities for well-being items and the well-being composite in samples A, B, and C

Interrater reliabilities	Sample		
	A	B	C
Descriptive statistics			
n_{subjects}	155	176	149
n_{raters}	51	71	100
n_{ratings}	483	610	392
Maximum number of raters	5	7	6
Mean \pm SD raters per subject	3.12 ± 0.57	3.47 ± 1.45	2.63 ± 1.03
$ICC(3,k)$			
Item 1: Moods	0.76	0.75	0.71
Item 2: Social	0.72	0.79	0.72
Item 3: Goals	0.74	0.81	0.50
Item 4: Be subject	0.74	0.68	0.65
Well-being	0.81	0.83	0.73

In this table, n_{subjects} indicates the number of subjects used in the analyses, n_{raters} indicates the number of raters used in the analyses, and n_{ratings} indicates the total number of ratings in the analyses.

