Supplemental Material

## Method

**Coding of individual task performance.** The present study was designed to examine physiological stress responses; however, we also explored whether performance on the individual task was impacted by the experimental manipulation. Three independent raters coded individual task performance on two dimensions: quality of ideas and quality of nonverbal aspects of pitch. For each 30-second segment of the individual performance task, coders (blind to hypotheses and experimental condition) rated (a) quality of ideas (taking into account how well the required elements of the pitch were answered, how well thought out and explained the elements were, and how creative the ideas were) and (b) quality of nonverbal aspects of pitch (taking into account fluency of speech, confidence, and any other feature of the pitch that contributed to a perception of effectiveness besides the words that were spoken) on 0 (*worst*) to 6 (*best*) scales. Coders had access to both audio and video, were only able to view the participant who was speaking when making their ratings, and overlapped with each other on 13% of the corpus of video recordings, showing excellent inter-rater reliability on quality of ideas (alpha = .90) and quality of nonverbal aspects of pitch (alpha = .91).

**Self-report measures assessed for use in a different manuscript.** After baseline physiological recordings, participants filled out eligibility confirmation (including pregnancy and pacemaker status), demographics, rated their affect, and completed individual difference measures regarding emotion regulation, power, status, personality traits, and communal orientation. After the collaborative work task, participants rated self and teammate affect, anticipatory stress, social perceptions of self and teammate, and connection to teammate. After the individual performance task, participants rated performance, self and teammate affect, social perceptions of self and teammate, connection to teammate, and perceptions of self and teammate emotion regulation.

**Results**

**Baseline physiological measures: PEP, CO, and TPR.** No baseline physiological differences were observed among the experimental conditions. For manipulated teammates, emotion regulation condition was not associated with differences in baseline PEP ($F(2,127) = 1.59$, $p = .21$), CO ($F(2,127) = 0.16$, $p = .85$) or TPR ($F(2,127) = 1.54$, $p = .22$). For non-manipulated teammates, emotion regulation condition was not associated with differences in baseline PEP ($F(2,127) = 0.06$, $p = .94$), CO ($F(2,131) = 1.14$, $p = .32$) or TPR ($F(2,131) = 1.75$, $p = .18$; see Table S1 for descriptive statistics).

**Gender moderation.** Gender did not moderate the effect of stress reappraisal on physiological responses. For manipulated teammates, gender does not moderate the effect of stress reappraisal on CO reactivity during collaborative work task ($F(2,119) = 1.20$, $p = .305$), CO reactivity during individual performance task ($F(2,121) = 0.12$, $p = .883$), TPR reactivity during collaborative work task ($F(2,118) = 0.48$, $p = .618$), or TPR reactivity during individual performance task ($F(2,107) = 2.52$, $p = .085$). For non-manipulated teammates, gender also did not moderate the effect of arousal reappraisal on CO reactivity during collaborative work task ($F(2,123) = 1.38$, $p = .256$), CO reactivity during individual performance task ($F(2,116) = 0.40$, $p = .670$), TPR reactivity during collaborative work task ($F(2, 121) = 2.51$, $p = .086$), or TPR reactivity during individual performance task ($F(2,108) = 0.22$, $p = .802$).

**Dyadic analyses.**

***PEP reactivity during collaborative work task.*** For dyads, emotion regulation condition did not influence PEP reactivity ($F(2,132) = 0.50$, $p = .607$).

***Table S1***. *Raw Score Descriptive Statistics for Physiological Variables*

| | | Manipulated teammates | | | | | | Non-manipulated teammates | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Control | | Suppression | | Reappraisal | | Control | | Suppression | | Reappraisal | |
| | | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Baseline recording | PEP | 107.19 | 14.78 | 106.42 | 14.18 | 111.46 | 13.31 | 107.66 | 12.66 | 106.90 | 14.60 | 105.88 | 11.50 |
| | CO | 7.72 | 1.90 | 7.45 | 2.24 | 7.56 | 2.42 | 7.72 | 2.24 | 7.37 | 1.90 | 7.03 | 2.23 |
| | TPR | 870.90 | 254.31 | 974.85 | 374.42 | 1007.78 | 463.07 | 928.59 | 348.98 | 957.35 | 276.37 | 1065.14 | 439.16 |
| Collaborative work task | PEP | 96.80 | 16.20 | 96.94 | 15.25 | 99.25 | 18.10 | 99.16 | 20.69 | 98.61 | 14.70 | 95.86 | 12.72 |
| | CO | 7.53 | 1.79 | 7.32 | 2.58 | 8.01 | 2.35 | 7.76 | 2.28 | 7.05 | 1.94 | 7.27 | 2.18 |
| | TPR | 1016.60 | 267.22 | 1152.56 | 478.64 | 1031.83 | 437.55 | 1039.84 | 415.17 | 1118.61 | 327.71 | 1060.39 | 378.92 |
| Individual performance task | PEP | 82.30 | 14.02 | 85.33 | 18.37 | 83.22 | 16.75 | 84.16 | 14.90 | 82.44 | 12.71 | 83.46 | 15.16 |
| | CO | 8.47 | 2.17 | 7.98 | 2.87 | 9.28 | 3.08 | 8.17 | 2.35 | 7.93 | 1.96 | 8.19 | 2.67 |
| | TPR | 1031.35 | 283.62 | 1191.63 | 424.85 | 934.59 | 367.08 | 1121.96 | 450.13 | 1149.55 | 331.89 | 1037.60 | 349.95 |

*Note*. PEP is measured in milliseconds; CO is measured in liters per minute; TPR is measured in dyne-seconds/$cm^5$.

***CO reactivity during collaborative work task***. Emotion regulation condition had an impact on CO reactivity ($F(2,132) = 9.96$, $p < .001$). Stress reappraisal dyads exhibited greater increases in CO from baseline ($M = 0.28$, $SD = 0.70$) relative to both the suppression ($M = -0.26$, $SD = 0.83$; $F(1,88) = 19.38$, $p < .001$) and control conditions ($M = -0.17$, $SD = 0.92$; $F(1,85) = 12.47$, $p < .001$).

***TPR reactivity during collaborative work task***. For dyads, emotion regulation condition impacted TPR reactivity ($F(2,127) = 24.91$, $p < .001$). The reappraisal condition produced significantly lower (more efficient) TPR relative to both the suppression ($F(1,86) = 44.50$, $p < .001$) and control conditions ($F(1,78) = 30.30$, $p < .001$), which did not differ from each other ($F(1,88) = 2.72$, $p = .103$).

***PEP reactivity during individual performance task***. For dyads, emotion regulation condition did not impact PEP reactivity ($F(2,127) = 1.09$, $p = .337$).

***CO reactivity during individual performance task***. For dyads, emotion regulation condition impacted CO reactivity ($F(2,116) = 10.44$, $p < .001$). The reappraisal condition produced significantly higher (more efficient) CO reactivity relative to both the suppression ($F(1,87) = 15.73$, $p < .001$) and control conditions ($F(1,71) = 14.87$, $p < .001$).

***TPR reactivity during individual performance task***. For dyads, emotion regulation condition impacted TPR reactivity ($F(2,113) = 32.90$, $p < .001$). The reappraisal condition produced significantly lower (more efficient) TPR reactivity relative to both the suppression ($F(1,77) = 50.43$, $p < .001$) and control conditions ($F(1,80) = 44.31$, $p < .001$), which did not differ from each other ($F(1,75) = 0.24$, $p = .620$).

**Interpersonal mechanisms of contagion.**

***Interaction analyses for affective displays.*** Similar to the results presented in the manuscript, and consistent with the notion that manipulated reappraisers' expressions of PA may have impacted non-manipulated teammates' physiology, we observed a significant condition (reappraisal vs. control+suppression combined) * manipulated teammates' collaborative task PA interaction on non-manipulated teammates' individual task CO, $F(1,105) = 15.75$, $p < .001$, 95% CI [0.76, 2.27], $\eta_p^2 = 0.13$. In the combined control+suppression conditions, manipulated teammates' PA was negatively related to non-manipulated teammates' CO, $b = -0.30$, $t(67) = -2.51$, $p = .015$.

Also similar to the main results, and consistent with the notion that manipulated reappraisers' expressions of NA may have impacted non-manipulated teammates' physiology, we observed a significant condition (reappraisal vs. control+suppression combined) * manipulated teammates' collaborative task NA interaction on non-manipulated teammates' individual task TPR, $F(1,101) = 9.01$, $p = .003$, 95% CI [103.67, 508.01], $\eta_p^2 = 0.08$. In the combined control+suppression conditions, manipulated teammates' NA was negatively related to non-manipulated teammates' TPR, $b = -0.26$, $t(64) = -2.15$, $p = .036$.

***Additional mechanism interaction analyses.*** We did not observe a significant condition (reappraisal vs. control) * variable interaction on non-manipulated teammates' CO for warmth ($F(1,69) = 0.16$, $p = .691$), competence ($F(1,69) = 1.16$, $p = .285$), or sense of connection ($F(1,69) = 0.03$, $p = .865$). Nor did we observe a significant condition (reappraisal vs. control) * variable interaction on non-manipulated teammates' TPR for warmth ($F(1,64) = 0.05$, $p = .819$), competence ($F(1,64) = 0.45$, $p = .507$), or sense of connection ($F(1,64) = 1.23$, $p = .272$).

**Classification analyses.** Past research has shown that physiological reactivity to stress may show two distinct profiles: challenge (higher CO reactivity and low-to-no TPR reactivity) and threat (low-to-high CO reactivity and high TPR reactivity). The present data are consistent with these patterns, but we can also use the data to assess multivariate empirical clusters of physiological reactivity. In other words, using untrained clustering algorithms, do the physiological reactivity indices spontaneously create indices of challenge and threat, and do the conditions differ by profile classification?

To investigate this, we used the k-means algorithm from the flexclust package (Leisch, 2006; v1.3-5 in *R* v3.5.1) on all the physiological reactivity indices available during the preparation period (CO, TPR, RSA, PEP, SV). First, we z-scored all the variables and removed subjects with missing data, leaving 232 complete observations. Then, we ran the k-means algorithm using Euclidean distance, checked the cluster centers (see Table S2), and assigned each participant to their respective clusters. Finally, we performed a logistic regression to test whether cluster assignment significantly differs by Condition. Cluster 1 is consistent with

**Table S2**
*k-means cluster solution (z-scored) for the collaborative task*

| | CO | TPR | PEP | RSA | SV |
|---|---|---|---|---|---|
| Cluster 1 | 0.765 | -0.672 | -0.177 | 0.005 | 0.547 |
| Cluster 2 | -0.751 | 0.622 | 0.193 | 0.065 | -0.569 |

physiological challenge because it exhibits increased cardiac output reactivity and lower peripheral resistance, whereas Cluster 2 shows the opposite pattern. Confidence in these clusters was confirmed via the Breckenridge (2000) procedure. (To ensure reproducibility of the results, all procedures with variability from randomness—i.e., random data partitioning, k-means algorithm fitting—used the seed 1001, chosen arbitrarily.)

The total sample was randomly split into halves (n1 = n2 = 116). N1 and n2 were used to train k-means algorithms with 2 clusters. The solution for n1 was used to predict the classification of subjects from n2, and agreement between the two algorithms was computed using the Adjusted Rand Index (ARI), returning ARI = .71. The procedure was repeated for the other half, returning ARI = .87. These data are dyadic, so we also assessed cluster reliability by splitting the data into manipulated and non-manipulated partner sets and followed the same procedure. ARI for agreement between manipulated partner's cluster and manipulated partner's cluster predicted by the other k-means fit is .90, and ARI for agreement between non-manipulated partner's cluster and cluster predicted by the other algorithm is .80, suggesting good reliability (see full results in Table S3).

With these reliability metrics and the *a priori* dichotomy between challenge and threat profiles, we proceeded and analyzed how condition affected the probability of cluster assignment to the challenge profile. This was done by using the classifications from the k-means algorithm fit to the entire data set as the response variable in logistic regression, with cluster 1 coded as 1 (challenge) and cluster 2 coded as 0 (threat), and Condition as the predictor variable in a hierarchical model to account for dyadic data. Condition significantly improved the model fit compared to an intercept-only model, $\chi^2(2) = 18.94$, $p < .001$. Individual comparisons showed that the reappraisal condition significantly increased the probability of being in the challenge cluster compared to control ($b = 1.23$, $SE = 0.37$, OR = 3.42, 95% $CI_{OR}$ [1.71, 7.60], $\chi^2 = 3.32$, $p < .001$) and compared to suppression ($b = 1.41$, $SE = 0.39$, OR = 4.11, 95% $CI_{OR}$ [2.03, 10.17], $\chi^2 = 15.81$, $p < .001$) (see full logistic regression results in Table S4).

Another avenue of investigation is whether the physiological signals can reliably differentiate the conditions. Given that the suppression and control conditions did not significantly differ on cardiac efficiency, it is likely that a classification algorithm would be unable to differentiate them. However, it should be relatively better at classifying reappraisers.

Using the caret package (Kuhn, 2015; v6.0-80) and e1071 package (Dimitriadou et al., 2008; v1.7-0), we performed linear discriminant analysis (LDA). First, we broke the data into the training set with manipulated teammates and test set with non-manipulated teammates. The LDA was trained using LOOCV and had an accuracy of .41, 95% CI [.32, .50]; this was not significantly better than the No Information Rate (NIR) of .34, $p = .116$. We compared the trained model to LDAs fit to two conditions at a time (see results in Table S5) and as expected, it seems the high rate of error was because of the overlap between control and suppression condition. The following models are assessing the fit of the trained LDA reporting two-tailed *p*-values (see confusion matrices in Table S6).

Although the physiological signals did not provide clear linear discriminants between suppression and control, they did afford some accuracy in separating the two from reappraisal. Nevertheless, testing the trained model on the non-manipulated teammates demonstrated a better fit than the no information rate: The accuracy of classification was .52 (95% CI [.43, .62]), which was better than the NIR of .35, $p < .001$. Further, the model successfully classified non-manipulated teammates of reappraisers (see Table S7 for confusion matrix): 29/40 (72.5%) of non-manipulated teammates of reappraisers were accurately classified based solely on their physiological profile, whereas only 24/77 (31.2%) of non-manipulated teammates of suppressors or controls were false positives inaccurately classified as non-manipulated reappraisers, $\chi^2 = 18.15$, $p < .001$.

**Table S3**

*Number of participants in each condition classified as Cluster 0 (threat) or Cluster 1 (challenge)*

| | Cluster 0 (Threat) N | Cluster 1 (Challenge) N | Odds Ratio (OR) (Challenge/Threat) | CI lower | CI upper |
|---|---|---|---|---|---|
| Control | 43 | 30 | 0.70 | 0.40 | 1.12 |
| Suppression | 50 | 30 | 0.07 | 0.34 | 0.94 |
| Reappraisal | 24 | 55 | 0.52 | 1.44 | 4.14 |

**Table S4**

*Logistic regression results*

| | *B* | *SE* | Wald $\chi^2$ | *p* | OR | CI lower | CI upper |
|---|---|---|---|---|---|---|---|
| Intercept (Control) | -0.37 | 0.25 | -1.49 | .138 | 0.69 | 0.40 | 1.12 |
| Suppression | -0.16 | 0.35 | -0.45 | .650 | 0.85 | 0.42 | 1.71 |
| Reappraisal | 1.23 | 0.37 | 3.32 | .001 | 3.42 | 1.71 | 7.60 |

**Table S5**

*LDA model trained to two conditions at a time*

| | Accuracy (Acc) | CI lower (Acc) | CI upper (Acc) | No Info Rate (NIR) | *p*-value (Acc > NIR) |
|---|---|---|---|---|---|
| Reappraisal vs. Control | .68 | .51 | .74 | .51 | .039 |
| Reappraisal vs. Suppression | .60 | .49 | .71 | .50 | .089 |
| Control vs. Suppression | .42 | .31 | .54 | .51 | .109 |

**Individual task performance.** Experimental condition did not significantly impact manipulated teammates' (*F*(2,128) = 1.13, *p* = .326) or non-manipulated teammates' (*F*(2,128) = 1.62, *p* = .202) quality of ideas (see Figure S1, Panel A). For manipulated teammates, the reappraisal condition (*M* = 4.77, *SD* = 0.88) did not produce significantly different quality of ideas relative to the suppression (*M* = 4.47, *SD* = 1.08; *t*(128) = 1.44, *p* = .152) or control conditions (*M* = 4.70, *SD* = 1.00; *t*(128) = 0.36, *p* = .721), which did not significantly differ from each other (*t*(128) = 1.08, *p* = .283). Similarly, for non-manipulated teammates, the reappraisal condition (*M* = 4.73, *SD* = 0.87) did not produce significantly different quality of ideas relative to the suppression (*M* = 4.29, *SD* = 1.29; *t*(128) = 1.80, *p* = .074) or control conditions (*M* = 4.51, *SD* = 1.14; *t*(128) = 0.91, *p* = .363), which did not significantly differ from each other (*t*(128) = 0.89, *p* = .374).

Similarly, experimental condition did not significantly impact manipulated teammates' (*F*(2,128) = 0.97, *p* = .384) or non-manipulated teammates' (*F*(2,128) = 0.69, *p* = .505) quality of nonverbal aspects of pitch (i.e., *how* the pitch was delivered; see Figure S1, Panel B). For manipulated teammates, the reappraisal condition (*M* = 4.56, *SD* = 0.91) did not produce significantly different quality of nonverbal aspects of pitch relative to the suppression (*M* = 4.28, *SD* = 1.04; *t*(128) = 1.31, *p* = .191) or control conditions (*M* = 4.50, *SD* = 1.05; *t*(128) = 0.28, *p* = .784), which did not significantly differ from each other (*t*(128) = 1.04, *p* = .302). For non-manipulated teammates, the reappraisal condition (*M* = 4.48, *SD* = 0.87) did not produce significantly different quality of nonverbal aspects of pitch relative to the suppression (*M* = 4.20, *SD* = 1.31; *t*(128) = 1.14, *p* = .257) or control conditions (*M* = 4.39, *SD* = 1.21; *t*(128) = 0.35, *p* = .728), which did not significantly differ from each other (*t*(128) = 0.80, *p* = .426).

**Table S6**

*Assessing model performance: Confusion matrix for actual and predicted number of manipulated teammates by condition*

| | Predicted Control | Predicted Reappraisal |
|---|---|---|
| Actual Control | 21 | 16 |
| Actual Reappraisal | 12 | 27 |
| | Predicted Suppression | Predicted Reappraisal |
| Actual Suppression | 21 | 18 |
| Actual Reappraisal | 13 | 26 |
| | Predicted Control | Predicted Suppression |
| Actual Control | 13 | 24 |
| Actual Suppression | 20 | 19 |

*Note*: Numbers in each cell are counts.

**Table S7**

*Testing model on non-manipulated teammates: Confusion table for predicted and actual number of non-manipulated teammates in all three conditions*

| | Predicted Control | Predicted Suppression | Predicted Reappraisal |
|---|---|---|---|
| Actual Control | 12 | 9 | 15 |
| Actual Suppression | 12 | 20 | 9 |
| Actual Reappraisal | 6 | 5 | 29 |

*Note*: Numbers in each cell are counts.

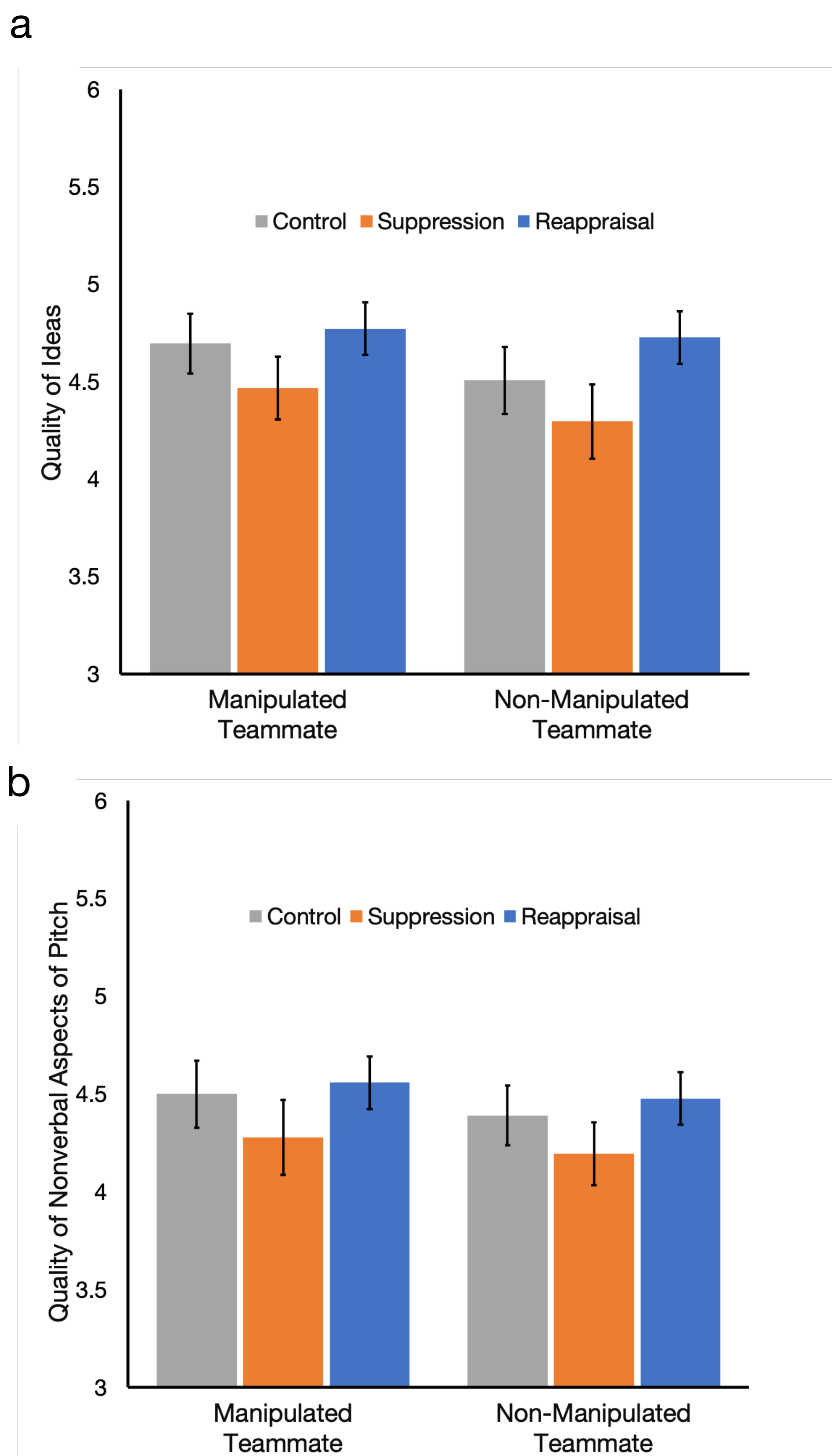


***Figure S1***. Coded performance (0-6 scale) during the individual task on quality of ideas (Panel A) and quality of nonverbal aspects of pitch (Panel B). Error bars represent one standard error.

## Supplemental Appendix

### Collaborative Work Task Instructions

*Part 1* of your pitch will focus on describing the UCSD [U of R] Bicycle. This 3-min portion of the presentation should cover the following details:

1) The Product
    a. Key features and price
    b. Design innovations
    c. Who and how big is the target market?
    d. What problem is the product solving for the customer? Why will they want to purchase the UCSD [U of R] Bicycle?

2) Present a timeline for all key milestones, including when the product will be profitable

*Part 2* of your pitch will focus on the marketing plan for the UCSD [U of R] Bicycle. This 3-min portion of the presentation should cover the following details:

1) The Marketing Plan
    a. What is your plan?
    b. Where/how will you advertise?
    c. How will you allocate your marketing budget?
    d. How will this product help promote the UCSD [U of R] brand?

2) Why are you the best team to lead this initiative?

Remember,
-Each of you should prepare to present both parts of the presentation; we will assign one of you to Part 1 at the start of the presentation period.
-You will give your presentation to trained evaluators, and your team's performance will be assessed based on the quality and originality of ideas, and how clearly and effectively you present your pitch.