Appendix C
Fs/QCA Procedures and Extra Analyses

1. Statistical Underpinnings of fs/QCA

Unlike other statistical procedures that rely on significance testing or variance explanation, fs/QCA relies on the principles of Boolean algebra and utilizes combinatorial logic to identify the causally relevant conditions that explain the phenomenon of interest (Ragin, 2008). Fs/QCA thus identifies sets of causal attributes that are common across all cases with regard to an outcome. In our study, the attributes are the cultural dimensions, a "case" is the country, and the outcome is the mean rho for the component-creativity relationship.

2. Data Calibration of Cultural Dimensions

For each cultural dimension we identified three breaking points: a point above which countries are considered to be very high on that dimension, a point below which countries are considered very low on that dimension, and a mid-point at which a country is between these extremes. For Hofstede's dimensions, the mid-point was set at 50 – the middle point of Hofstede's scale (1980) that represents the theoretical and empirical cutoff between high and low scores of a variable. The high-point was set at 75, and the low-point was set at 25 – two points previously used by Hofstede himself to distinguish between "moderately high" and "very high" scores (Hoftede, Hofstede, & Minkov, 2010). For cultural tightness, the mid-point was set at 6, the average value across countries included in the original study (Gelfand et al., 2006), with the high-point set at 8 and the low point set at 4.

3. Core vs. Peripheral Conditions

The distinction between core and peripheral conditions is realized by removing 'easy' (i.e., assumptions that are consistent with empirical and theoretical knowledge) and 'difficult' counterfactuals (i.e., assumptions that are consistent with empirical evidence but not with theoretical knowledge – see Ragin, 2008, for a thorough discussion). Core conditions are considered more decisive and definitive causal ingredients because they remain part of the solution even after the inclusion of both easy and difficult counterfactuals. Peripheral conditions are instead robust to the inclusion of easy counterfactuals, but are "stripped away" from the solution "if the researcher is willing to make assumptions that are at odds with existing substantive and theoretical knowledge" (Ragin & Fiss, 2008: 204; see Fiss, 2011, and Misangyi et al., 2017, for more detailed explanations).

4. Supplementary Analyses and Robustness Checks for fs/QCA

We performed various robustness analyses to ensure the stability of our configurations. First, we used a frequency cutoff of two instead of one, meaning that we require that at least two countries per configuration should be present before a solution is considered in the subsequent analysis. Second, we re-ran the analysis after removing from the dataset all countries for which there was only one effect size in the dataset, as this effect size could potentially be non-representative of the entire country.

Third, we changed the threshold values for the outcome condition. In other words, we moved the upper threshold and the lower threshold to see if results are robust when countries are reclassified given the outcome. We did this by replacing standard deviations with weighted standard deviations. In short, the first two changes altered the set of countries considered in the analysis, whereas the third changed the configurational fit score of each configuration. We found

that every configuration reported in Table 8a is still part of the solution in at least two of the three alternative specifications used for our robustness analysis. This provides strong evidence of the robustness of the solutions and of our findings (Greckhamer et al., 2018). For Table 8b, we see that every configuration is present in at least one of our alternative specifications. While this provides some evidence of robustness, this set of solutions seems to be less stable (Greckhamer et al., 2018).

5. Exploratory fs/QCA analysis – Substitution patterns

Our findings also reveal tradeoffs/substitution patterns between specific cultural dimensions within bundles that strengthen the components-creativity relationship: when one dimension is present, the other is absent, and vice-versa. These tradeoffs could not be identified through our meta-analysis, and are worth exploring because they provide further insights into the cultural bundles configurations that strengthen the components-creativity relationship.

When a bundle includes cultural looseness/does not include cultural tightness, individualism becomes a necessary, core condition to strengthen the effect of domain-relevant skills on creativity. This finding suggests a tradeoff between the presence of tightness and the presence of individualism in order for domain-relevant skills to have a stronger effect on creativity. Comparing our fs/QCA findings with the truth table for domain-relevant skills corroborates this idea: while some bundles that include tightness and three values emphasizing domain-relevant skills strengthen the effect of domain-relevant skills on creativity (e.g., Mexico and Saudi Arabia – see Table 8a, panel 1, configuration 3), others that present the same configuration and are present in our sample do not (e.g., Germany and Italy). The difference between these two sets of cultural bundles is that in the former the value that according to our

predictions does not promote the use of domain-relevant skills is high power distance, whereas in the latter it is individualism.

The same tradeoff can be observed for task motivation: tightness and individualism are never simultaneously included in the bundles that strengthen the effect of this component (see Table 8a, Panel 3).

6. Calculation of Minimum-Fit Scores

Fs/QCA analysis produces configurations (i.e. cultural bundles) that show stronger and weaker component-creativity relationships. Based on these configurations, we computed the minimum fit score for each country to create a cultural bundle variable that could subsequently be used in our variance-known, three-level hierarchical linear meta-analytical regressions.

The minimum fit score measures the degree to which each country belongs to the configurations that strengthen the component-creativity relationships. In other words, this score (ranging from 0 to 1) indicates how far or close the configuration of cultural dimensions of a particular country is from an "ideal" bundle found to be conducive to a stronger component-creativity relationship. The closer a country score is to 1, the closer the country is to have the same conditions (i.e., the same cultural dimensions) of the "best" configuration.

For illustration, consider the following hypothetical example. Let uss assume that a configuration that strengthens a given component-creativity relationship is one that includes Cultural Tightness·•High Uncertainty Avoidance•Masculinity•High Power Distance•Individualism. If a country scores 1 for cultural tightness, 1 for uncertainty avoidance, 1 for masculinity, 1 for power distance and 1 for individualism, the fit score would be equal to 1 – a perfect match. If instead ta country scores less than 1 for any two dimensions (e.g., uncertainty

avoidance and masculinity – respectively 0.9 and 0.8), the minimum fit score would be equal to the lowest score registered – i.e., equal to 0.8 for the whole configuration.

As it is possible to have more than one high-performance configuration for each component, when this was the case we took the highest minimum fit score for a country across the multiple configurations and as the "cultural bundles fit" variable used in the analysis. Thus, if we identified three solutions, and a country had minimum fit scores of 0.9, 0.8, and 0.7 respectively, we took 0.9 as the value for the "cultural bundles fit" variable.