Supplementary Methods

* 1. *Literature searches*

Searches of the literature were carried out in Web of Science, Scopus, ProceedingsFirst, and Google Scholar from October to December 2022, using the terms “pollution”, “contaminant”, “noise”, “light” or “turbidity” with the terms “social\* behaviour”, “sociability” or “group cohesion”, and the term “fish” (Google Scholar only) (Figure 1). Searches were limited to English language studies and those within the field of zoology. Papers used included those published in peer-reviewed journal articles as well as unpublished theses (5.5%). Relevant papers that were missed by literature searches were also identified from citations (Alfonso *et al.* 2020; Fahlman *et al.* 2021; Salahinejad *et al.* 2022). Sources were only included for analyses if shoaling or schooling were examined in relation to anthropogenic contaminants and pollutants, including chemical contaminants, noise, light, and turbidity. Analysis was therefore limited to species that are known to exhibit shoaling or schooling behaviour. Where species were not specified as shoaling or schooling in relevant papers e.g., in trials of individual sociability or aggregation, searches of the literature were conducted to ascertain whether each species displays shoaling or schooling behaviour. As this review focused on effects on fish species, studies of chemical compounds not designed to test the effects of toxicants in aquatic systems, such as use of zebrafish *Danio rerio* as pharmacological models, were excluded (Wiprich *et al.* 2020). Studies were excluded if they did not directly quantify either sociability or group cohesion, meaning studies of sexual behaviour, aggression, dominance hierarchies, migration, social learning, and social recognition were excluded. Interactions with non-anthropogenic stressors (e.g., predation cues) were also excluded. Interactions with temperature were included, as fish are ectothermic and therefore particularly dependent on environmental temperature for determining rates of a variety of relevant physiological processes; evidence exists of additive effects of temperature with contaminants on fish physiology (Morgan *et al.* 2001). Papers using numerical models (theoretical studies) were excluded, as were conspecific vs. heterospecific preference tests. Papers studying “aggregation” were not included (Pitcher 1983), with one exception where “aggregation” was specifically used as a proxy for wild shoaling in a shoaling species (Clark 2022; Clark and Ioannou 2022). All papers were considered independently by one author of the current study (I.C.T.), and then blind reviewed by one of two other authors (D.C. or A.M.). Papers that were disagreed upon were discussed and included or excluded by consensus, based on the above-mentioned considerations.

* 1. *Data collection*

Following paper inclusion, data on the following variables were collected for each paper. In full factorial studies, which was the case for all but two (Holcombe *et al.* 1980; Porseryd *et al.* 2017) papers investigating interactions between multiple contaminants and/or pollutants, the effects of individual- and multiple-stressor trials were recorded separately.

1. Effect size of each treatment on cohesion or sociability, relative to the control. As studies used different metrics to assess sociability and cohesion, direct comparison using the units given in studies was not possible. Therefore, any change in behaviour was converted to a standardised measure, effect size, that could then be used in models to compare results among studies. Effect size (Hedges’ *g*, a standardize difference of means between groups) of treatments (i.e. contaminant or pollutant exposure) were calculated for 141 studies and 1816 treatments, as some or all treatments from 24 studies (Currie 2021; David *et al.* 2015; Ehlman *et al.* 2019; Fahlman *et al.* 2021; Fewtrell and McCauley 2012; Holcombe *et al.* 1980; Iwama *et al.* 1986; Kaplan *et al.* 2013; Kastelein *et al.* 2017; Lafoux *et al.* 2023; Lal *et al.* 1984; McCallum *et al.* 2017; Mehdi *et al.* 2022; Neo *et al.* 2014; Neo *et al.* 2015; Patch *et al.* 2022; Petitjean *et al.* 2021; Rene Eslava-Mocha *et al.* 2019; Renick 2014; Rojas *et al.* 2023; van der Knaap et al. 2022; Ward *et al.* 2008; Weis and Weis 1974; Williams 1989) were excluded where some or all data needed to calculate effect size was not available. Effect sizes were calculated as a standardized measure of change in sociability or cohesion, as various metrics may be used to measure each. Effect sizes were transformed (multiplied by 1 or -1) according to the metric used to measure cohesion or sociability, so that positive effect sizes indicate an increase in sociability or cohesion and negative effect sizes a decrease.

Effect size of each treatment (Cohen’s *d*) on each metric of cohesion or sociability were calculated according to the formula:

$$Cd=\frac{\overbar{x}\_{1}-\overbar{x}\_{2}}{PSD}$$

Where Cd is Cohen’s *d*, $\overbar{x}$ is the mean of each group, and PSD is the pooled standard deviation of both groups, calculated as:

$$PSD=\sqrt{\frac{(N^{1}-1)SD\_{1}^{2}+(N^{2}-1)SD\_{2}^{2}}{N^{1}+N^{2}-2}}$$

Where N is the number of replicates of each group, and SD is the standard deviation of each group. Cohen’s d values were converted to Hedges’ *g* values (sample-size corrected effect size) using the formula:

$$Hg=(1-\frac{3}{\left(4×\left(N^{1}+N^{2}\right)\right)-9})×Cd$$

Where standard error rather than standard deviation was reported in a paper, standard deviation was calculated according to:

$$σ=SE\sqrt{n}$$

Where the standard error of the difference between groups was given (but not SE of both groups separately), pooled SE was converted to SD which was then used as pooled SD. Where confidence intervals were reported, standard deviation was calculated according to:

$$σ=\frac{\sqrt{N}×(upper limit-lower limit)}{3.92} $$

Where median and quartile values were reported, these were used in place of mean and standard deviation according to (Greco *et al.* 2015). Effect size was calculated for each treatment level, e.g. where trials were carried out at three temperature treatments, the effect size for each of these treatments was calculated relative to the control. Where multiple measurements were carried out within each treatment e.g., across time scale, means at each sampling point were recorded and used to calculate the overall mean and SD. Raw data was used to calculate means and standard deviations if these were not reported in the paper. Where exact values were not reported and raw data was not available, data was extracted from graphs using WebPlotDigitizer version 4.6 (Rohatgi 2022). For forest plot construction, mean and standard error of effect size for each study was calculated from effect sizes calculated for all treatment levels.

1. Focal species used in each paper. Species names differed in some studies from the modern accepted scientific name; in these cases, the modern accepted name was used according to the World Register of Marine Species database (WoRMS 2023). Species habitat was also recorded, as marine, freshwater, brackish, or any combination of these, based on the WoRMS database.
2. Class of contaminant or pollutant to which fish were exposed. Among non-chemical pollutants, noise, light, and turbidity were included. Classes of chemical contaminants were organic (including medicines, organically derived pesticides, oil and oil derivatives including microplastics, and androgens), inorganic (metals and metal salts), and mixed. Mixed contaminants included both contaminants of unspecified makeup, such as industrial or agricultural effluent, and specific combinations of contaminants used in lab settings.
3. Treatment levels of variables of interest. For trials with more than one treatment level, effect size was calculated for each treatment level relative to the control, or relative to the highest or lowest level (depending on the variable) for wild population comparisons. For combined stressors, effect size for all trials with non-control values of both or all stressors were calculated relative to trials using control values for both or all stressors. Where treatments were introduced partway through/comparative to different control treatments, changes were compared by subtracting Hedges *g* values for control before vs. after from treatment before vs. after trials (Morris 2008). Where this was not possible due to data presentation, values from after treatment introduction were compared to the control. Actual (not nominal) treatment levels were recorded where these were given. For fluctuating treatments, the effect size was calculated between control and the maximum treatment level. As different contaminants may produce effects at different concentrations, variation in concentration was high for many substances. Treatment levels were therefore converted to z scores using the formula:

$$z= \frac{x- \overbar{x}}{σ}$$

Where $x$ is the value, $\overbar{x}$ is the mean, and $σ$ is the standard deviation of the mean.

1. Group sizes. This was defined as the number of fish used per group trial in group cohesion studies. Where group size varied, the minimum number of fish per group in a treatment was used. It was not possible to use an average number due to the way group numbers were reported in some papers.
2. Sex of fish(es) used in each experiment. This was recorded as unsexed where fish were not separated by sex, or where reported data was pooled for both sexes.
3. Number of replicates. Number of replicates was recorded as number of groups per treatment. Where the number of replicates per treatment varied e.g., between control and treatment groups, the value for treatment groups was recorded (or the average value if this was reported). If trials were repeated on the same groups, n was recorded as the number of groups. In effect size calculations, number of groups was also used, unless only one replicate was present in which case number of fish per group was used as the calculations used do not allow a sample size of one.
4. Period over which fishes were exposed to each treatment. Exposure time was recorded in days, with appropriate fractions for hours and minutes (e.g., 1h = 1/24 = 0.0417 days). Where a range of exposure times was reported (e.g., 3-4 days), the minimum exposure period was recorded. Exposure times include acclimation/gradual reduction or increase times. Where individuals were raised under treatment conditions, age at assessment of social behaviour was used. Recovery period (if applicable) following exposure but prior to assessment of social behaviour was also recorded.
5. Exposure type. While related to exposure period, this allowed us to separate wild and lab studies, and direct exposure from developmental or parental exposure. Categories of exposure type were uncontrolled (including populations exposed in the wild, and farmed fish exposed in a natural setting), developmental, parental, acute, or acclimated. Developmental studies were defined as those where fish were exposed during the egg and/or larval phase, then allowed to grow into adults before behavioural trials were conducted. Parental exposure was defined as parents of tested fish being exposed, but not the tested fish themselves. Acute exposure was defined as up to seven days direct exposure of adult or juvenile fish, while acclimated exposure was defined as seven days or more direct exposure of adult or juvenile fish.
6. Metrics used to assess cohesion or sociability. In the majority of studies, individual sociability was quantified using some variant of the three-chamber test (Xia *et al.* 2010) or shoal preference test (Al Marshoudi *et al.* 2023; Salahinejad *et al.* 2022), while group cohesion was quantified in group shoaling trials (Lanzarin *et al.* 2020; Santos *et al.* 2021). Cohesion within groups of fish was commonly measured as inter-individual distance , nearest-neighbour distance, shoal polarity, or mean distance from the centre of the shoal (Table 1). Individual sociability was often measured as the frequency of approach to a conspecific shoal, time spent near a conspecific shoal, latency to approach a conspecific shoal, or distance from a conspecific shoal (Table 1). Other studies measured sociability as time spent away from the conspecific shoal, or sociability score, a function of time and distance from a conspecific shoal (Table 1). Various studies used time spent in different tank “zones” to measure social behaviour in shoal choice tests, however exact study designs varied slightly. Therefore, “time away from social zone” was used to refer to trials in which experimental arenas were divided into only two zones, and was only used if it was the only variable reported, being directly proportional to time in social zone. “Time in far zone” was used in studies in which arenas were divided into three or more zones, and refers to farthest zone from the stimulus shoal. Where zone times were statistically compared e.g., near vs. far zone, but time in each zone was not compared between treatments, results were recorded as statistically significant if a difference in the relationship between treatments was found e.g., if fish spent more time near the shoal in the control treatment, but had no preference in other treatments.
7. Reported statistical significance (or lack thereof) of the effect of each treatment were recorded for each paper. Statistical significance was recorded as *p* < 0.05. Where only the overall relationship between a continuous stressor and cohesion or sociability was tested, individual treatment levels were marked as significant (or not) according to the result of the overall relationship.
8. Country where each study was carried out (geographical location).
	1. *Statistical analysis*

A metanalysis was then carried out in R using the MCMCglmm package (Hadfield 2010) to run phylogenetically-adjusted linear models for each contaminant or pollutant class. Classes were modelled separately, in order to reduce model complexity and allow interactions among explanatory variables for each contaminant or pollutant class to be thoroughly explored. To account for phylogenetic relatedness among fish species, a phylogenetic tree was constructed including all species, using the R packages phytools (Revell 2012) and phylobase (Hackathon *et al.* 2020) which allow trees to be constructed from the Tree of Life database (Tree of Life Web Project 2023). This tree was then incorporated into a dataset with effect size and variable data using mulTree package (Guillerme and Healy 2020), which combines a data table with a phylogenetic tree object to form a data list. The resultant response variable in these models was effect size, while the explanatory variables were level of the relevant stressor (e.g. contaminant concentration or pollutant level for each treatment), exposure period, number of replicates, group size, exposure type (acute, acclimated, repeated, developmental, or uncontrolled), and metric (group cohesion or sociability). For studies of wild populations, i.e. where exposure period and group size were not known, 100 days and group size of 10 individuals (arbitrarily chosen values corresponding to long-term exposure and high group size relative to lab studies) were used to allow inclusion of wild studies in models. The interaction effect between contaminant concentration or pollutant level and metric was also included where multiple concentrations were present within each metric. In the case of light pollution, an interaction between the form of light exposure and treatment was included in order to separate effects of luminance from different wavelengths. While metric was likely somewhat collinear with group size, interactions between these variables were not included as the majority of variables had only one value for sociability trial group size (one). Specimen, referring to different treatments both within and among papers utilising the same species, was also included as a random effect variable to account for the effects of different trials, both within and across studies, using the same species. Species was also included as a random effect variable, as is required to quantify phylogenetic effects in MCMCglmm models. Sex of fish(es) used in each trial was also included as a random effect variable. Multi-stressor interactions were not modelled due to the small number of studies that investigated stressor combinations. Lambda (λ) value, and indicator of the effect of phylogenetic relatedness among species on effect size, for phylogenetic signal was calculated using the phytools package. Graphs and figures were plotted using the ggplot (Wickham 2016), ggraph (Pedersen 2022), and ggtree (Yu *et al.* 2017) packages in R.